The Thesis Committee for Paul Amador
Certifies that this is the approved version of the following thesis:

Novel Methods for Exploration and Engineering of Regulatory ncRNA in Bacteria

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Lydia M. Contreras

Hal S. Alper

Novel Methods for Exploration and Engineering of Regulatory ncRNA in Bacteria

by

Paul Amador, BS

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

December 2016

Acknowledgements

I would like to thank Dr. Lydia Contreras and my fellow lab members for their help and support throughout my graduate studies. My growth and development as a researcher was bolstered by the continuous encouragement and direction of Dr. Contreras. I am also particularly grateful to Steve Sowa for his mentorship and guidance.

The work presented in this thesis results from collaborations with Jordan Villa and Abigail Leistra. Their contributions were substantial to the design and experimentation of the work presented herein. Many of the large experiments conducted were aided by the volunteer efforts of lab members, particularly, Chen Tsai and Mark Sherman.

Finally, I would like to thank my family for instilling a passion for science and providing unconditional support.

Abstract

Novel Methods for Exploration and Engineering of Regulatory ncRNA in Bacteria

Paul Amador, MA

The University of Texas at Austin, 2016

Supervisor: Lydia M. Contreras

The need for sustainable resources has spurred the establishment of microorganisms as platforms of chemical production. These "microbial factories" are engineered to maximize production of valuable chemicals. Essential to the engineering of these production strains, is the ability to control and regulate gene expression. The engineering of native and synthetic pathways of production relies on the ability to control the levels of intermediate and end products, as well as the mechanisms that impose control and induce gene expression. Traditionally, this involves the targeted deletion and overexpression of specific genes of interest within the genome. Of late, however, there has been interest in exploring the regulatory capacity of RNA for biotechnology. Discovery of regulatory elements, such as riboswitches and sponge RNAs, has advanced the capacity and utility of RNA for bioengineering. In my work, I have developed an *in vivo* screen for detecting RNA elements that are responsive to stress in the radiation-resistant bacteria, *Deinococcus radiodurans*. Investigations of the response and

regulatory capacity of ncRNA, specifically 5' UTRs, are especially valuable in this organism as radiation induces stress similar to that of oxidation and aging. Notably, this work has yielded the discovery of an mRNA-based regulon, previously thought to act only at the promoter level to regulate multiple genes associated with radiation and desiccation response. The expanded RNA-based model I present for this regulon serves to validate the advantages of RNA-level regulation for rapid and robust responses that make such RNA regulatory elements valuable for biotechnological applications. Moreover, my research into RNA regulation involved the characterization and engineering of the regulatory ncRNA, *csrB*, that, in concert with the regulatory protein CsrA, controls the expression of hundreds of genes throughout gamma proteobacteria. Specifically, I utilized an *in vivo* assay to generate an accessibility profile for *csrB* (for which little structural information exists in *E.coli*) to determine sites that are likely binding the regulatory protein, CsrA. This accessibility profile informed the rational engineering of *csrB* to alter affinity for binding CsrA dimers and ultimately "tune" the regulatory capacity of the Csr global regulatory system for producing complex phenotypes desirable for "microbial factories".

Table of Contents

List of Tables

List of Figures

Chapter 1: RNA as an Emerging Tool for Bioengineering

For centuries humans have used microorganisms to produce commodities such as wine and cheese. And as a result of our ancestor's epicurean pursuits, the science of manipulating microbes flourished. Indeed, the lessons learned from fermenting food products were expounded to yield medicines and antibiotics. As the value of these products became apparent, so did the need for a stronger scientific investment. Consequently, many of the biological mechanisms behind the production of these compounds were progressively elucidated. Whereas, initially, microbes were merely selected and implemented to produce the compound of interest, nowadays, a microbe's biology can be wholly altered or engineered to improve or expand its ability to produce biological commodities. Moreover, a heightened environmental and ecological awareness have sounded the horn for more sustainable means of chemical and energy production. As such, the field of bioengineering has developed and evolved to enhance the control that can be imposed on biology to serve essential or novel functions.

Early efforts to improve or alter the synthetic capacities of microbes involved artificial selection strategies to enrich specific desirable phenotypes. Over time, this approach for directed evolution has become much more sophisticated and even automated(1, 2). Moreover, uncovering the phenotype to genotype link has been facilitated by the advent of next-generation sequencing technologies. These advances have revealed the dynamic nature between the genome, transcriptome, and proteome and uncovered networks of genes that contribute to these complex phenotypes (3). As a result, the primary methods of engineering organisms have relied on turning genes "ON" or "OFF" by deleting genes or overexpressing them. Traditionally, this practice involves modifications at the genomic level, and can include imposing genes under non-native

promoter control. However, this DNA-centric approach neglects other important loci of gene expression that occur at the post-transcriptional level. In fact, some of the more successful engineering strategies optimize gene expression at both the transcriptional and translational level (4).

The guiding principles that inform post-transcriptional, RNA-level engineering are generally derived from nature. Complex regulatory systems have been discovered in prokaryotes and eukaryotes that generally serve to modulate expression as a response to stress or other stimuli. Regulating gene expression at the RNA level has many advantages over DNA level regulation; since transcripts are only one step away (translation) from their effector function, responses are much faster. In addition, multiple copies of mRNA produce a much more robust response and the generally short half-life of transcripts mean the responses can be much more dynamic and temporally controlled. One of the more striking discoveries is the CRISPR (clustered regularly interspaced short palindromic repeats) system that essentially acts as an adaptive immune response in certain prokaryotes. This system recalls nucleic acid sequences of historical pathogens and inhibits expression by employing complementary guide RNAs and the accessory protein, Cas9, to inhibit expression (5). Similarly, other inhibitory responses have been observed that utilize anti-sense RNAs (asRNA) in complex with proteins to affect gene expression, such as miRNA-RISC and sRNA-HFQ (6, 7). Recently, regulatory elements have been found on mRNA that involve the untranslated regions (UTR) (8–10). Some examples, such as riboswitches, rely on structured RNA in the UTR that undergoes conformational changes when bound to a ligand to affect expression or termination. These are often auto-regulatory mechanisms that contribute to maintenance of proper levels of proteins or metabolites (11). The UTRs can also be important sites of regulation by global regulatory systems, as in the carbon storage regulator (Csr) system, where an RNA sequence motif

recruits a regulatory protein (12). The mechanisms of these regulatory systems have been exploited for bioengineering and synthetic biology tools, as illustrated below, Figure 1.1. Currently, many of the sRNA engineering strategies depend on base-pairing to target DNA or mRNA and use engineered protein complexes to elicit a change in expression (13, 14). These efforts take full of advantage of the rapid and robust nature of RNA regulation and offer efficient control of expression for single or specific targets. However, recent studies have highlighted the value of engineering global regulation to produce complex phenotypes better suited for biosynthesis of commodity chemicals (15, 16). Current RNA-based engineering efforts may be constrained to targeting unique genes and do not proffer a facile route towards engineering global regulation.

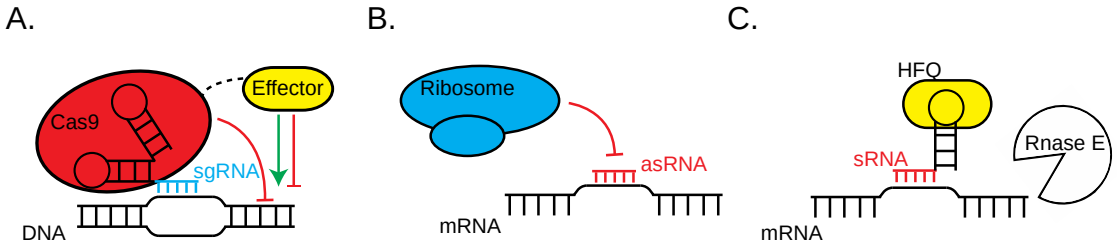

Figure 1.1: *RNA-based tools for synthetic biology*. A, engineered CRISPR/Cas9 tools utilize single guide RNAs (sgRNA) to target genes at the genome or transcript level and can inhibit by Cas9 or include effector proteins (fused to Cas9) that impose control of expression. B, asRNA can be designed to bind specific transcripts and abrogate translation. C, sRNAs designed to bind specific transcripts can be fused to HFQ-binding RNA, which recruits Rnase E to induce degradation of target.

In this work, my aims were to: first, explore the RNA-based regulation of the extremophile bacteria, *Deinococcus radiodurans*, at 5' UTRs under radiation stress, and second, to alter global regulation in *Escherichia coli* by the rational engineering of the non-coding RNA (ncRNA), *csrB*. As current RNA-based engineering strategies have

been inspired by nature, an initial exploratory work was conducted to elucidate novel 5'UTR-based regulatory elements in a valuable model organism. In brief, an *in vivo* fluorescent reporter was developed to test 5'UTRs that respond to ionizing radiation stress. The specific aim was to reveal 5'UTRs that may be valuable for responding to a biologically relevant stress (ionizing radiation/oxidation) and to discern the mechanism, be it a riboswitch or other novel function. This work has uncovered a potentially novel function for a recently discovered regulatory system in *D.radiodurans*. The radiation and desiccation response (RDR) regulon is a set of radiation responsive genes that exhibit a specific sequence motif RDR motif (RDRM), and thought to be controlled at the promoter level (17). However, my studies reveal that gene activation may also be occurring at the post-transcriptional level to produce a more efficient stress response. My more applied work involved the rational engineering of *csrB*, a major component of the Csr system present throughout gamma proteobacteria (12). The Csr system involves the protein, CsrA, which regulates a large set of genes to alter carbon metabolism and other functions. The ncRNA, *csrB,* sequesters CsrA to antagonize its regulatory effects (18). This work involved an *in vivo* assay to determine sites along *csrB* that could be binding CsrA. The CsrA-binding sites were systematically altered to produce a collection of *csrB* species that exhibit a gradient of CsrA regulation and potentially represent "fine tuning" of the complex Csr-associated phenotypes, which may be valuable for improved biosynthesis. The results of my thesis work serve to fortify the importance of RNA in regulation in biological systems. Additionally, the engineering of the ncRNA, *csrB,* represent early efforts to regulate genes en masse or engineer entire regulatory systems. The novel approaches utilized in this work may be adaptable for other organisms and advance more grandiose engineering efforts to enhance our control over biology.

Chapter 2: Genome-wide Screen for Regulatory 5'UTRs

INTRODUCTION

Life has evolved and adapted to inhabit the vast majority of our planet. Microorganisms, such as bacteria, are especially adaptable and have been found to survive in hostile conditions such as extreme temperatures, high salinity, desiccation, and other stressors (19). The demands imposed by such environments have elicited critical adaptations for survival. Central to surviving extreme conditions is an organism's ability to respond to external stresses by exerting tight control over gene expression. Indeed, regulation of gene expression can occur at three levels: transcriptional (at the DNA level), post-transcriptional (at the RNA level), and post-translational (at the protein level). Many of the well-characterized stress responses in bacteria act at the DNA level to bind promoter sequences to repress or activate transcription (20, 21). However, large-scale investigations into genome, transcriptome, and proteome dynamics, facilitated by increased accessibility to next-generation sequencing technology, have uncovered many novel regulatory patterns for cells under stress (22–26). Notably, regulation has been observed at the post-transcriptional and post-translational level.

The need for timely and precise control of gene expression is similarly imperative for extremophile bacteria such as, *Deinococcus radiodurans*. This Gram-positive bacterium serves as a model organism for studying extreme stress tolerance due to the exceptional resistance to ionizing radiation (IR), desiccation, and oxidative stress that has been observed. This bacterium has evolved the ability to resist thousands more times the IR dosage lethal to humans, upwards of 15,000 Grey (Gy) (27–29). The effects of IR are particularly damaging as it can cause double-stranded DNA breaks (DSB) (28). Moreover, IR can generate reactive oxygen species (ROS) that oxidize nucleic acids and proteins and affect their natural processes (28). To confront these challenges,

D.radiodurans has evolved to harbor multiple copies of its genome. This helps to resist DSBs and facilitates the repair process, as there are multiple templates for recombination to occur (27, 30). This organism also exhibits efficient means of clearing damaged proteins and oligonucleotides by degrading them, which, in turn, increases protein turnover for facilitating expression of new proteins that may contribute to repair (28). To combat the accumulation of harmful ROS, *D.radiodurans* also produces anti-oxidants such as deinoxanthin, which contributes to its red pigment, and novel Mn-Fe complexes that have been observed to scavenge ROS (27, 31, 32).

However, many of the genes involved in these repair processes are conserved in many other radiation-sensitive organisms (33). Moreover, *D.radiodurans,* has been observed to sustain as much damage as radiation-sensitive organisms and appears to have no special adaptations for "shielding" its genome or proteins from IR damage (27, 31). Thus, it is likely that how these genes are regulated and coordinated plays a critical role in post-IR recovery. Recently, many investigators have sought to elucidate a novel regulatory system in *D.radiodurans* known as the radiation and desiccation response (RDR) regulon. This regulatory system is thought to activate the expression of approximately 29 genes associated with radiation and oxidation stress responses, such as DNA repair enzymes (17). The regulation is mediated by the protein homodimer, DdrO, which binds at a specific 17 base pair motif, known as the radiation and desiccation response motif (RDRM), which is present upstream of the coding sequences of these RDR genes. Upon exposure to radiation stress, the protease, IrrE, is activated to remove DdrO and allow for expression of the RDR genes (17, 34). Furthermore, a quorum-sensing system has been identified in *D.radiodurans*, that increases expression of enzymes that synthesize N-acyl-L-homoserine lactones (AHL) under oxidative (H_2O_2) stress, which serve to activate stress response genes (35). While quorum sensing is

present in many other bacteria, this is the first example of quorum sensing observed in *D.radiodurans*. Similarly, riboswitches, such as the FMN and glmS riboswitches, that are present in other organisms, have also been identified in *D.radiodurans* (36, 37). Recent investigations have also examined the repertoire of small non-coding RNA (sRNA) that are differentially expressed under IR stress and could contribute to regulation of stress response genes (38). The prospect of RNA-based regulation in *D.radiodurans* is especially intriguing, as RNA molecules are much shorter and may be less vulnerable to DSBs (27, 38). Regulation at the RNA level also requires less biological processing to respond to stress. Many sRNAs, function by binding complementary DNA or RNA to affect gene expression, and only require transcription to be effective (8, 39). Other RNA regulatory elements, such as the aforementioned riboswitches, are present in the untranslated regions (UTR) of mRNA and can be especially efficient for eliciting a stress response (11, 40). The UTRs can also exhibit other regulatory functions, such as binding sites for regulatory proteins or RNA (40, 41). Regulation via UTR possesses many ideal qualities for stress responses: local regulation of abundant mRNAs can provide a much stronger activation of expression than regulation at the promoter level, UTR regulation is only one step removed from a protein effector function to provide quick responses, the shorter half-life of mRNAs provides a more temporally dynamic response, and finally, there is a reduced metabolic load for organisms in recovery since there is no need for transcription.

In this work, previously generated *D. radiodurans* transcriptome data was utilized along with structural predictions and conservation analysis to identify 5'UTRs that regulate gene expression under ionizing radiation (38). This bioinformatics approach ultimately resulted in 41 5'UTR candidates that were verified by RT-PCR and 5'RACE. A novel *in vivo* fluorescence reporter expressing a translational fusion of the 5'UTR to a

codon-optimized green fluorescent protein (GFP) was developed to test the candidate 5' UTRs for responses to IR. In addition, candidate 5' UTRs were also tested under H_2O_2 oxidation stress. The most prominent result has been discovery that the GyraseA (dr1913) 5' UTR may regulate expression as a response to increasing IR. The gene (dr1913) is part of the RDR regulon and thought to be regulated at the transcriptional level via repression at the promoter by IrrE (17, 34, 42, 43). Notably, this work suggests the GyraseA 5' UTR may also be regulated at the post-transcriptional mRNA level. Further experiments revealed potential binding of IrrE to the GyraseA 5'UTR. Thus, this work suggests novel mRNA-level regulation for the RDR regulatory system. This expanded model of RDR functionality, serves to fortify the potential contribution RNA-level regulatory systems make for improving stress responses.

METHODOLOGY

Plasmids and strains

All plasmids and strains are included in Table 2.1. Wild-type *D. radiodurans* R1 (ATCC13939) served as the parent strain utilized for all the *Deinococcus radiodurans* test strains and mutants. The KatA- strain (44) was kindly provided by Dr. Michael Daly. The 5'UTR disruption mutants were constructed in this study by collaborator Roland Saldanha with the Air Force Research Labs. The Δ *IrrE* strain was gifted by Dr. Pascale Servant (34). The plasmids used for testing differential expression of 5'UTR-GFP fusions (including the positive control, pRadGro-GFP) were synthesized from the pRadGro vector (45) by Genscript®. Upon sequencing, we determined a derivation of this plasmid from the published results, as noted in Figure 2.1. The theophylline riboswitch construct was a gift from Dr. Nancy Kelley-Loughnane. Moreover, the pRadGro-ThRS-GFP

plasmid was produced by cloning the theophylline riboswitch upstream of GFP with SacI and ApaI cut sites. Protein expression was conducted in *E.coli* BL21(DE3) cells and all cloning was conducted in *E.coli* DH5α. All *D.radiodurans* strains were cultured in TGY (1% Tryptone, 0.5% yeast extract; and 0.1% glucose) media (BD Difco) at 30˚C. All *E.coli* strains were cultured in Luria-Bertani (LB) media (BD Difco) and grown at 37˚C. The strains and plasmids used for these experiments are catalogued in Table 2.1.

Plasmids and strains	**Description**	**Source**	**Plasmid Map**
Plasmids			
pRadGro	Shuttle vector for *E.coli* and *D.rad*; Chl^R, Amp^R	*Misra et al. 2006*	Figure 2.1, A
pRadGro-ThRS-GFP	pRadGro-GFP with theophylline riboswitch upstream of GFP	This Study	Figure 2.1, B
pRadGro-5'UTR-GFP	pRadGro-GFP reporter plasmid with candidate 5'UTR sequences inserted upstream of GFP for translational control	This Study	Figure 2.1, C
pRadGro-GFP	pRadGro plasmid expressing green fluorescent protein	This Study	Figure 2.1, D
pET28a-His-DdrO	Expression vector used for purifying DdrO protein	This Study	
Strains			
D.radiodurans			
R1	Wild-type *D.radiodurans* (ATCC 13939)		
KatA-	Deletion of catalase KatA Dr_1998; Kan^R	*Markillie et al. 1999* (KKW7003)	
DR349ΔUTR	R1 strain with ATP-dependent protease UTR region disrupted	This Study	
DR1279ΔUTR	R1 strain with Mn Superoxide dismutase UTR region disrupted	This Study	
DR1857ΔUTR	R1 strain with OHRP UTR region disrupted	This Study	
DR1913ΔUTR	R1 strain with GyraseA UTR region disrupted	This Study	
Δ IrrE	Δ IrrE Ω kan (R1 strain with knockout of *IrrE*)	*Devigne et al. 2015*	
E.coli		(GY 14127)	
BL21(DE3)	Strain used for protein expression and purification	Lab Stock	
DH5α	Strain used for all cloning	Lab Stock	

Table 2.1: *Plasmids and strains*

Figure 2.1: *Plasmid maps of pRadGro constructs.* A, map for the pRadGro parent plasmid with deviation from published sequence highlighted (Red) in text box. B, plasmid with theophylline riboswitch construct and color-matched sequence. C, plasmid with 5'UTR construct and color-matched sequence. D, plasmid with GFP under native promoter control.

Transformation of Deinococcus radiodurans

Transformation of *Deinococcus radiodurans* was performed based on a previous procedure with minimal modifications (46). *Deinococcus radiodurans R1* grown to late-log phase (OD_{600} ~1) were mixed with 30 mM $CaCl_2$ and 10% glycerol to make competent and stored at -80˚C. Competent cells were incubated with 1 µg of plasmid DNA on ice for 30 minutes, followed by incubation at 30˚C for 1 hour. Transformed cells were then diluted 1:4 with fresh TGY media in a new tube and grown at 30˚C for 18 hours. After incubation cells were pelleted at 3000 rpm for 5 minutes and resuspended in 150 µL TGY for plating on to TGY plates with the appropriate antibiotic. Plates were then incubated for two days at 30˚C.

Reverse-transcription PCR (RT-PCR)

As an initial means of experimentally confirming the presence of predicted 5'UTR regions in their corresponding mRNAs, RT-PCR analysis was conducted. Total RNA was extracted from WT *D.radiodurans R1* at mid-log phase (OD_{600} 0.6-0.8) using vigorous bead beating and TriZol as previously described (38). The extracted RNA was then treated with DnaseI (New England Biolabs) for 1 hour at 37˚C and denatured at 65˚C. Complementary DNA was obtained using random hexamer oligos (N_6) (though some reactions were troubleshooted with gene-specific primers and Superscript III Reverse Transcription Kit (Invitrogen) according to manufacturer's protocols. The cDNA was then subjected to PCR with primers designed to yield amplicons of just the coding region and the coding region plus the 5'UTR (if present). The PCR products were run on 1% agarose gels and visualized with EZ-vision dye (AMRESCO) on a ChemiDoc XRS+ imager (Bio Rad).

5' Rapid Amplification of cDNA Ends (RACE)

For determining the transcriptional start sites (TSS) of 5' UTRs, rapid amplification of cDNA ends (RACE) was conducted. Specifically, the FirstChoice®RLM-RACE kit (Ambion) was utilized per manufacturer protocol. In brief, total RNA was extracted from WT *D.radiodurans R1* and 10 µg ligated to the 5' RACE Adaptor (provided in kit) using T4 RNA ligase at 37°C for one hour. The RNA-5' RACE Adaptor product was then reverse-transcribed, as before, using the provided M-MLV reverse transcriptase enzyme and random decamer primers (N_{12}) at 42°C for one hour. The 5' UTR cDNA was then PCR amplified and sequenced (Sanger method by UT core facilities) to determine the TSS.

Conservation analysis using LocARNA

As a means of identifying functional 5' UTR candidates, conservation of the candidate 5' UTRs in *D.radiodurans R1* were compared to the sequences of 45 organisms including all members of the most closely-related *Deinococcus* and *Thermus* phylum (28) with annotated genomes, and more distantly related organisms including, *Mycobacterium leprae, Bacillus subtilis, Escherichia coli, Saccharomyces cerevisiae.* The full list of organisms is presented in the results as a phylogenetic map indicating relative evolutionary distance**.** Candidate 5' UTRs were subjected to BLAST (NCBI) for identifying conservation levels in any of the aforementioned 45 organisms and any sequences identified with an E value less than 10^{-4} were utilized for conservation analysis and structure prediction with LocARNA (47–49). LocARNA utilizes ClustalW to align multiple sequences and generate consensus structures based on conserved sequences and base pairing probability.

Fluorescence cytometry

A Benton-Dickinson FACScalibur with a 488 nm argon laser and 530 nm FL1 logarithmic amplifier was utilized for observing differential expression of green fluorescent protein (GFP) in our 5'UTR test strains under various stressors. For each sample, 100 µL was pelleted and resuspended in 1 mL filter-sterilized 1X Phosphate-buffered Saline (PBS) in polyethylene cytometer tubes (Falcon). As we were unable to identify any prior protocols for assaying fluorescence in *Deinococcus* with a cytometer, some parameters had to be established. Initially, *Deinococcus* cells had to be gated, or isolated, based on FSC v SSC to gate for the viable *D. radiodurans* population (see results). More importantly, the sequence for GFP was codon-optimized for expression in *D.radiodurans* (see results). CellQuest Pro (BD) software was used to attain the median fluorescence of each triplicate cell populations. Induction, or activation of GFP expression, was assessed by determining the average value for the ratios of paired replicates of the treated samples over the sham. To determine significant induction values, a one-tailed student's t-test was performed (Microsoft Excel) comparing the induction of samples to the induction of the empty vector (EV, pRadGro), followed by another t-test comparing the induction of samples to the induction of the pRadGro-GFP vector (GFP) under GroES promoter control, which should not exhibit true induction(45).

Theophylline riboswitch activity assay

Establishment of the fluorescence screen for the 5'UTR-GFP fusion candidates was done using the well-characterized synthetic theophylline riboswitch (ThRS) (50). *D. radiodurans R1* containing the 5'UTR-GFP plasmid were grown to mid-log (OD_{600} 0.6-

0.8) and induced with 2 mM theophylline (control with DMSO). Fluorescence was measured via fluorescence cytometry two hours after induction as described above.

High-dose irradiations (1-15 kGy)

Samples (5 mL volume cultures) of *D. radiodurans R1* were harvested at mid-log phase (OD_{600} 0.6-0.8) and double-sealed in polyethylene bags (2-oz. Whirl-Pak® Bags; Nasco) that were subsequently placed inside larger polyethylene specimen bags (7-oz. Whirl-Pak® Bags; Nasco). Samples were then refrigerated in coolers with dry ice and transported to the irradiation facility. Acute doses of ionizing radiation (1 kGy-14 kGy) were conducted with a 10-MeV, 18-kW linear accelerator (LINAC) β-ray source at the Texas A&M National Center for Electron Beam Research, Texas A&M, College Station, Texas. The staff at the center operated the LINAC source for proper calibration and dosimetry of the ionizing beta radiation for all samples. The samples were irradiated on wet ice to keep the cell growth static. Shortly after irradiation, samples were transferred to 15 mL conical tubes and diluted two-fold in fresh media. The samples were allowed to recover in incubators at 30˚C for 2 hours prior to fluorescence cytometry (as described above).

IR Survival Assay

To obtain the comparable percentage of survival for each of the IR dosages tested (0, 1, 5, 10, 15 kGy) samples of *D. radiodurans R1* were grown and irradiated as previously described. However, samples were not permitted to recover and were instead

kept static on dry ice until plating. Samples were plated at various dilutions (from 10^{-3}-10^{-7}) on TGY agar plates and incubated at 30˚C for two days. Colonies were then counted and normalized to the counts at 0 kGy. Additionally, spot plates were done for each dilution.

Low-dose irradiations (100-500 Gy)

Samples were prepared as before for the high-dose irradiations in double-sealed polyethylene bags. For the low-dose irradiations (100-500 Gy) a dual source Cs-137 gamma irradiator was utilized. The samples were placed in the center of the irradiation chamber and dry ice was placed along the sides of the chamber to keep samples at 4˚C (ice was replenished as necessary). The samples were exposed at a rate of approximately 47.1 Gy/hour and recovered as before in fresh TGY media for 2 hours. Samples were then analyzed using fluorescence cytometry as described above.

Hydrogen Peroxide Stress Assay

Samples (5 mL volume cultures) of *D. radiodurans* strain KatA- were grown to OD_{600} 0.6-0.8 for treatment with 15 mM H_2O_2. The stress concentration was determined based on literature concentration ranges (36). Samples were incubated with 15 mM H_2O_2 for 30 min at 4˚C. Following treatment, samples were recovered with fresh media for a 2-fold dilution and incubated for 2 hours at 30˚C. After recovery, 100 µL sample was resuspended in 1 mL PBS for analysis by fluorescence cytometry as described above. Survival assays of WT and KatA- strains treated with H_2O_2 were performed as described above for IR survival.

Northern Blotting

Northern blotting was used to evaluate mRNA differential expression as a result of irradiation. Total RNA was prepared from *D. radiodurans R1* cells exposed to 0 kGy, 1 kGy, and 10 kGy as described in the RT-PCR protocol. We used a 8% polyacrylamide gel (National Diagnostics) for total RNA electrophoresis (under denaturing conditions), and a total of 5 μg of RNA was loaded onto each lane for sampling. PhiX174 DNA/HinfI ladder (catalog number E3511; Promega), radiolabeled with γ-32P by T4 polynucleotide kinase (catalog number M0236S; New England BioLabs), was used as a size marker. The separated RNAs were transferred onto a positively charged membrane (Hybond N+, catalog number RPN119B; GE Life Sciences) and cross-linked with 254 nm UV light. PerfectHyb Plus hybridization buffer (catalog number H7033; Sigma-Aldrich) was used for probe hybridizations over 3 h of incubation at 42°C. Radioactivity was recorded by phosphor storage imaging (Typhoon; GE). The oligonucleotide probes were designed to have a complementary sequence toward the target RNA and radiolabeled with γ-32P by T4 polynucleotide kinase (catalog number M0236S; New England BioLabs). The sequences for the GFP and tRNA24 (loading control) probes were: 5' TCTTGAAGAAGTCGTGCT 3' and 5' GGTTTCCCGCTCGATTTTGAGTC 3', respectively.

Western Blotting

D.radiodurans protein lysate was obtained from cells grown to an OD_{600} of ~0.8 (10 mL cultures) and treated with the appropriate IR stress as described above. The cell pellet was stored at -80˚C post-irradiation until processing. The pellet was resuspended in 500 μL of PBS and lysed with 10 mg/mL lysozyme (Sigma-Aldrich) overnight at room temperature. Resuspension was further lysed using a probe sonicator voltage output of

approximately 9V for 1 min for six bursts with 10 min rest on ice in between to prevent overheating and denaturing of proteins. Following sonication, sample was then centrifuged at 15,000 rpm to pellet the cellular debris and insoluble protein. Soluble protein lysate was obtained from the supernatant and quantified using a Bradford assay (utilizing coomassie protein assay reagent by Thermo Scientific). Western blotting was performed with minimum modifications from previously published protocols (51, 52). SDS-PAGE was performed using a 12% bis-acrylamide gel and loaded with 4 μg total protein of each protein sample in an appropriate amount of denaturing SDS loading buffer (125 mM Tris (pH 6.8), 25% glycerol, 2% SDS, 0.01% bromophenol blue, and 0.5% β-mercaptoethanol). All loading volumes were made constant by diluting protein samples with appropriate amounts of sterile water and samples were boiled before loading for protein denaturation. Gels were run in duplicate to allow for one to be stained by SYPRO-ruby by manufactures' protocol (Invitrogen) as a loading control as the specific levels of typical house keeping proteins were not assumed to be constant under irradiation. Following SDS-PAGE, we transferred the gel to a 0.2 μm nitrocellulose membrane (Bio-Rad catalog #1620112) by electroblotting at 17 V for 35 min using trans-blot semi-dry transfer cell (Bio-Rad). Detection of GFP was achieved by using an anti-GFP antibody (Roche/11814460001) at a 1:1000 dilution and an anti-mouse-HRP conjugate (Promega/W4021) at a 1:2500 dilution. All antibodies were diluted in 1% non-fat milk (in Tris buffered saline (TBS): 20 mM Tris, 500 mM NaCl), and membranes were blocked overnight at 4 °C with 5% nonfat milk (in TBS) to minimize nonspecific binding. Anti-GFP was added first to membrane and incubated at room temperature for 1 hr. After incubation, the membrane was extensively washed with TTBS (TBS with 0.05% Tween 20) and then TBS. Membranes were then incubated with the Anti-Mouse-HRP conjugate for 1 hr. After washing, chemiluminescent detection using Bio-Rad's clarity

western ECL substrate provided images of bands of interest (GFP). Quantitation of western bands was done utilizing Total Lab Quant band volume detection protocol, which was normalized to a stable protein band in the SYPRO-ruby gel.

Mass Spectrometry

Protein lysate was obtained as described above and measured using Bradford assay. Equal amounts of protein were run on a 12% SDS-PAGE gel. Gel bands between 60 and 100 kDa (as determined by ladder) were cut and in-gel trypsin-digested based on previous published protocols (53). Briefly, cut gel bans were dehydrated with 100% acetonitrile then reduced with 10mM DTT for 30 min. Then the gel was alkylated with 50 mM idoacetamide at room temperature for 30 min. Gel bands were washed with 100 mM ammonium bicarbonate solution then dehydrated again with 100% acetonitrile. The dehydrated gel was then digested with 10 ng/μL trypsin (in 50 mM ammonium bicarbonate) overnight at 37˚C. Protein was then extracted from the gel using 5% formic acid and 1:2 (v/v) 5% formic acid: acetonitrile. Before running on mass spectrometer, protein samples were dried in a SpeedVac, then washed with 0.1% formic acid followed by a 0.5% TFA wash in ZipTip before elution into Elution Buffer (67% ACN; 32.8% Water; 0.2% TFA). Samples were then injected into a Thermo Orbitrap Elite hybrid linear ion trap FT-MS with Dionex 3000 nanospray UPLC and run for 1hr per sample. Peak areas were calculated using the Skyline Proteomics software (MacCoss Lab Software) and normalized to total ion intensity.

Gel Mobility Shift Assay (GMSA)

GMSA was performed with modest modifications from previous procedures (42). First, the 5’ UTR region of *GyrA* (DR1913) and the DNA positive control (promoter

region of *GyrA* previously shown to bind DdrO) were amplified by PCR (primers listed in Table 2.2) from WT *D.radiodurans R1*. The purified PCR product of the 5' UTR was utilized as the template for subsequent *in vitro* transcription with the MEGAscript T7 kit (Ambion) per manufacturer instructions. DdrO protein (DR2574) was expressed from a pET28a vector containing a N-terminal hexahistidine tag with a thrombin cleavage site. This vector (pET28a-His-DdrO) was produced by restriction cloning using NheI and BamHI cutsites and the primers described in Table 2.2.

Primer	**Description**	**Sequence (5' → 3')**
GyrA_RDRM_RNA_F	Forward primer for amplifying *gyrA* 5' UTR template for *in vitro* transcription	GAATTCAATTAATACGACTCACTATAGGGAGAAGTTGAGGAGAGCCGTTTT
GyrA_RDRM_RNA_R	Reverse primer for amplifying *gyrA* 5' UTR template for *in vitro* transcription	GTAGTTGATGAAATTGGTCTTGACTTCGC
Pdr1913 F	Forward primer for amplifying *gyrA* promoter region from Wang *et a*l. 2015	TTTTCTTCGTCTTTTCCGGTAAACTG
Pdr1913 R	Reverse primer for amplifying *gyrA* promoter region from Wang *et a*l. 2015	GGCACTCGCCGGGGGATA
DdrO _F_(NheI)	Forward primer for amplifying *ddrO* and cloning into pET28a vector for protein expression	GGAATTGCTAGCATGACATTGAAACTGCACGAACGACTTCGT
DdrO _R_(BamHI)	Reverse primer for amplifying *ddrO* and cloning into pET28a vector for protein expression	ATTCCGGATCCTCAGTTCAGGATGCGTTTGAGATGCAG

Table 2.2: *Primers used for GMSA*

E. coli BL21(DE3) cells transformed with this plasmid were grown to exponential phase in LB media containing 30 μg/ml kanamycin where 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) was added to begin protein expression. Expression cultures were incubated at 37˚C for 4 hours before cells were harvested by centrifugation. Cells were lysed prior to protein purification by resuspension in lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole, 5mM $MgCl_2$, pH 8) and sonication (10 V for 10 seconds six times, with 30 s on ice between sonications). Cell debris was removed by centrifugation (4,000 rpm at 4˚C for 30min). The supernatant was then purified for DdrO by a Ni-NTA resin column (Qiagen) using three column-volume washes with wash buffer (50mM NaH_2PO_4, 300mM NaCl, 5mM $MgCl_2$, 20mM Imidazole, pH 8) to remove impurities, followed by a single-column volume elution using elution buffer (50mM NaH_2PO_4, 300mM NaCl, 5mM $MgCl_2$, 250mM Imidazole, pH 8). Purified DdrO was then concentrated using Amicon Ultra-15 centrafugal filter with utracel-30 membrane into storage buffer (50% Glycerol and 100mM Tris/HCl pH 8.0). The hexahistidine tag was removed by thrombin cleavage using Thrombin CleanCleave™ kit according to manufacture's protocol (Sigma-Aldrich) overnight at 24˚C. Cleaved His-tag was removed by Ni-NTA column by centrifugation and removing the supernatant. Protein concentration was then measured by Bradford assay and SDS-PAGE gel to confirm purification.

DNA and RNA samples were dephosphorylated using calf intestine phosphatase (CIP)(NEB) before radiolabeling to improve phosphorylation at 37˚C for 1 hr. ^{32}P radiolabeling was done using T4 PNK incubated at 37˚C for an hour. Phenol/chloroform RNA extractions were done between both the CIP and PNK to remove the reaction enzymes and buffer. Sample reactions containing excess radiolabeled DNA or radiolabeled RNA and protein at 0, 8, 16, 32 μM were incubated in binding buffer (20

mM Tris-HCl pH8, 150 mM NaCl, and 5 mM $MgCl_2$) at 30˚C for 20 min before loading on a 12% TB-PAGE and run for 4 hours at 140V. Bands were visualized on a phosphoimager (Typhoon FLA 9500; GE).

RESULTS

Systematic analysis of RNA-sequencing data coupled with bioinformatics approaches reveals novel 5'UTR candidates that could contribute to stress response regulation in Deinococcus radiodurans. A bioinformatics approach was developed for this work to identify general 5'UTR regions that may contribute to post-IR gene regulation in *D. radiodurans*. An initial set of 417 potential 5'UTR candidates was identified by visual inspection of previously published RNA-sequencing data of *D. radiodurans* (38) (Figure 2.2, A)**.** The selected regions are among putative and annotated open reading frames in the sequenced genome of wild-type *D. radiodurans R1* at the 5' end of their adjacent coding region. All potential candidate regions were further filtered by three main parameters: UTR length, detected transcript counts, and previously reported differential expression of the adjacent gene under irradiation; specifically, candidates shorter than 35 base pairs were eliminated, as this represents the length of the shortest known 5'UTR regulatory element (54). Similarly, candidates that had transcript counts comparable to their adjacent coding sequence were ranked more favorably. Finally, the candidate list was cross-referenced with all previously published post-IR differential gene expression data (55–59). The candidates that exhibited some degree of differential expression were prioritized based on the rationale that differentially expressed proteins could be regulated by their associated 5'UTR under irradiation. Furthermore, putative 5'UTR regions associated with well-characterized metabolic enzymes were also

prioritized as these have been shown in the literature to be highly regulated by 5'UTR regions (40, 60). Any possible 5'UTR that ranked exceptionally in any of these parameters was added to the list of candidates independently of the other parameters. Notably, this analysis also aimed to rule out potential 3'UTRs (of an adjacent upstream coding region), noncoding intergenic RNAs, or coding spacers on polycistronic mRNAs. Additionally, any 5'UTRs corresponding to tRNA synthetases or ribosomal proteins were discarded (Figure 2.2, B). The top 100 candidates (Table 2.3) were then tested for contiguous expression and transcription start site (TSS) by either RT-PCR or 5'RACE(primers listed in Table 2.5) (depending on which method worked more efficiently for each candidate) **(**Figure 2.2, C**)**. RT-PCR utilized two sets of primers (primers listed in Table 2.4); the first set (primers X and Z) aimed to amplify a longer mRNA region containing the hypothetical 5'UTR and the second set of primers (primers Y and Z) aimed to amplify the known adjacent coding mRNA. These primers were designed arbitrarily to yield an amplicon of a convenient length (100-800 nucleotides) for visualization by agarose gel electrophoresis. A negative control (lacking the RT polymerase) was used to control for potential genomic DNA contamination; Using the mapped sequences for all 5'UTRs of interest, the RT-PCR and RNA-seq data, we were able to designate specific UTR sequences for 41 of our 5' UTR candidates that were then selected for further study, Table 2.6.

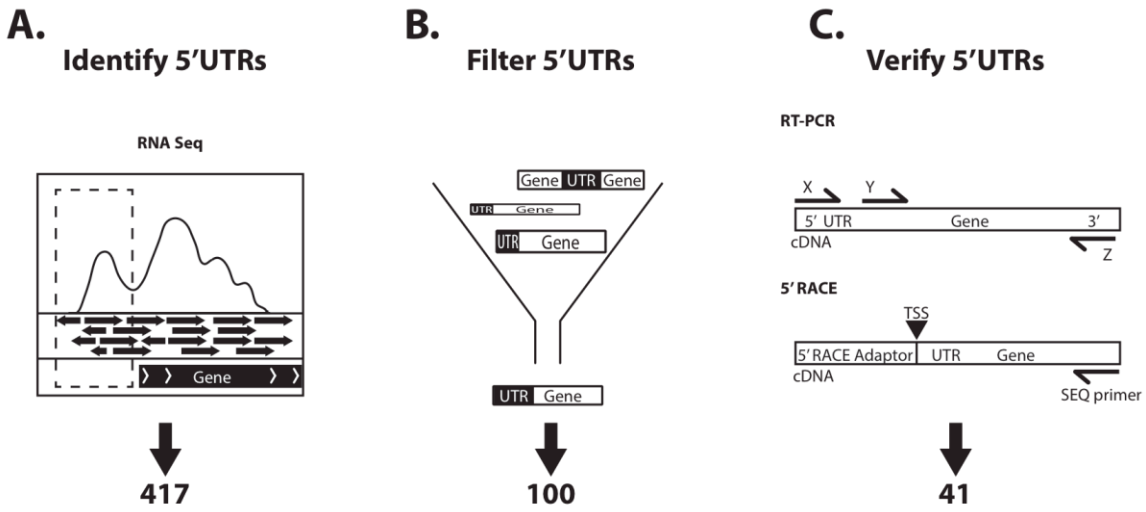

Figure 2.2: *Selecting Candidate 5'UTRs*. Process of selecting 5' UTR candidates included; A, inspecting RNA sequencing data for potential 5' UTRs upstream of annotated coding regions yielding 417 candidates, B, Potential candidates were further filtered based on size, location, and counts yielding 100 candidates, C, Top candidates verified by RT-PCR or 5'RACE yielding 41 final 5' UTR candidates.

Gene #	Gene Name	Length	Counts (approx)
29	2-oxo acid dehydrogenase, E1 component subunit alpha	55	45
35	adenylosuccinate synthetase	132	17
129	dnaK molecular chaperone DnaK	67	865
148	valS valyl-tRNA synthetase	89	105
153	riboflavin-specific deaminase/FMN switch	186	85
154	riboflavin synthase subunit alpha	67	66
183	glutamate synthase large subunit	99	18
294	putative nicotinate phosphoribosyltransferase	62	17
299	deoxyguanosine kinase/deoxyadenosine kinase subunit	56	115
302	glucosamine--fructose-6-phosphate aminotransferase	280	110
303	pgm phosphoglucomutase	67	90
325	malate dehydrogenase	40	25
349	ATP-dependent protease LA	110	245
353	ribonuclease	446	500
378	TetR family transcriptional regulator	311	180
433	beta-lactamase	62	30
464	maltooligosyltrehalose trehalohydrolase	287	50
551	glutaryl-CoA dehydrogenase	124	40
555	glutamate-1-semialdehyde aminotransferase	141	75
561	maltose ABC transporter periplasmic maltose-binding protein	217	170
562	maltose ABC transporter permease	58	110
599	aminoglycoside N3-acetyltransferase	134	580
606	groES co-chaperonin GroES	83	250
607	groEL chaperonin GroEL	77	100
640	methionine adenosyltransferase	79	60
674	argininosuccinate synthase	193	20
678	argininosuccinate lyase	153	45
788	branched-chain amino acid ABC transporter periplasmic amino acid-binding protein	82	240
813	1-pyrroline-5-carboxylate dehydrogenase	278	60
861	phytoene dehydrogenase	86	60
868	purH bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	71	35
873	O-acetylhomoserine (thiol)-lyase	136	700
899	ribonuclease H	202	16
932	polyprenyl synthase	70	60
951	sdhB succinate dehydrogenase iron-sulfur subunit	71	90
952	succinate dehydrogenase flavoprotein subunit	73	90
954	succinate dehydrogenase, cytochrome subunit	115	65
997	CRP/FNR family transcriptional regulator	96	55
1019	glycerol-3-phosphate dehydrogenase	125	90
1031	(S)-2-hydroxy-acid oxidase	60	45
1055	aspC aspartyl-tRNA synthetase	158	260
1081	TetR family transcriptional regulator	123	70
1084	methylmalonyl-CoA mutase		
1093	tryptophanyl-tRNA synthetase II	169	100
1105	DNA repair protein RadA	108	15
1120	butyrate kinase		
1122	prephenate dehydrogenase	69	50
1123	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	175	600
1147	chorismate mutase/prephenate dehydratase	73	45

Table 2.3: *100 Top Candidate 5'UTRs*

1156	transcriptional regulator	132	65
1160	uricase	72	60
1225	mannosyltransferase	65	100
1247	sucC succinyl-CoA synthetase subunit beta	86	2100
1276	seryl-tRNA synthetase	63	40
1279	Mn family superoxide dismutase	65	300
1304	spermidine/putrescine ABC transporter permease	128	20
1376	hypoxanthine-guanine phosphoribosyltransferase	73	40
1384	TetR family transcriptional regulator	123	20
1433	methionyl-tRNA synthetase	252	25
1456	ribose-phosphate pyrophosphokinase	60	35
1482	2-isopropylmalate synthase	134	125
1487	enoyl-CoA hydratase/3,2-trans-enoyl-CoA isomerase/3-hydroxyacyl-CoA dehydrogenase	97	50
1516	acetolactate synthase large subunit	217	200
1532	transcription-repair coupling factor	232	25
1573	pyrG CTP synthase	780	20
1613	ArsR family transcriptional regulator	152	160
1617	xanthine phosphoribosyltransferase	123	40
1698	tRNA-dihydrouridine synthase A	86	85
1718	glutamate dehydrogenase	82	25
1766	anthranilate synthase component II	134	300
1778	3-isopropylmalate dehydratase large subunit	170	40
1817	phospho-2-dehydro-3-deoxyheptonate aldolase	79	65
1906	L-lactate permease	140	17
1907	fumarate reductase-like protein	85	17
2008	aspartate-semialdehyde dehydrogenase	78	50
2087	translation initiation factor IF-3	130	80
2174	leucyl-tRNA synthetase	76	800
2206	citrate lyase subunit beta	117	60
2208	lactoylglutathione lyase	176	100
2217	terE tellurium resistance protein TerE	75	30
2221	tellurium resistance protein TerD	152	30
2225	tellurium resistance protein TerD	112	20
2259	transcriptional regulator	84	70
2285	A/G-specific adenine glycosylase	103	80
2303	chloramphenicol acetyltransferase	91	20
2306	MerR family transcriptional regulator	202	40
2354	pheS phenylalanyl-tRNA synthetase subunit alpha	102	115
2361	acyl-CoA dehydrogenase	165	65
2396	TetR family transcriptional regulator	97	450
2438	endonuclease III	179	70
2477	3-hydroxyacyl-CoA dehydrogenase		
2496	UDP-N-acetylmuramoylalanine--D-glutamiate ligase	99	20
2547	glutamyl-tRNA reductase	132	700
2568	argS arginyl-tRNA synthetase	127	1800
2588	iron ABC transporter periplasmic substrate-binding protein	124	250
2630	thyA thymidylate synthase	339	100
2635	pyruvate kinase	114	75
2637	eno phosphopyruvate hydratase	121	55
1662/1663	tdh L-threonine 3-dehydrogenase	492	120
2428/2429	bifunctional nicotinamide mononucleotide adenylyltransferase/ADP-ribose pyrophosphatase	563	50

Table 2.3: *100 Top Candidate 5'UTRs, cont.*

Gene #	Gene Name	Primer X (5' UTR fwd)	Primer Y (mRNA fwd)	Primer Z (mRNA rev)	Length of mRNA amplicon	Length of amplicon with 5' UTR	5'UTR Present
	Random Hexamer	NNNNNN	NNNNNN	NNNNNN			
29	2-oxo acid dehydrogenase, E1 component subunit alpha	ACGCTCGTTAGCTAGACA GATTTCA	GTGCCGATGTTCACGCTG ATT	TCGTTGCCGTCCACGTA GAAG	401	783	Yes
35	adenylosuccinate synthetase	GCAGCCGATAAGCTAAAC TGAATTACG	CCCGAGAAGTTCATCGAG GAAC	AGTTGAGGTCGAGGAGC GTG	436	776	No
129	dnaK molecular chaperone DnaK	CAGTTCAAGACTGCTGCC CTAAG	GACGAAGTGAAGGAAGA AGCGG	CGCAGGTCGAAGTTGTG TTCC	506	792	Yes
148	valS valyl-tRNA synthetase	CAACTTTCCACTTACCTT AGACCCTTTGA	TCGACAACACGCTCATCG ACA	CGATGCGGATTTCGCCA GTTTC	501	748	Yes
153	riboflavin-specific deaminase/FMN switch	GAAAGAAGGAGGAGACGC TCAG	TACGGAACATCCGCCTGA C	ATTCGGGCTGGTTCGGC CCAGG	82	137	Yes
154	riboflavin synthase subunit alpha	CTTAGACGTTAAGACCTT TTGACCCTT	GCCAGGAAGTCAATCTGG AGC	TCGTCGTCCACGAGGAT GAC	498	776	Yes
183	glutamate synthase large subunit	AACAACGGCCCCGCATAA AG	CTCAAGATTCTGGAAAAC CTCGACCA	GCAGCAGACCCTTATAG ATGATGGTC	502	791	No
294	putative nicotinate phosphoribosyltransferase	CTCTGCCCCTGCCTCATT G	CAAGCTGCCGCTGGAAAT CC	CGTTGATGGCGTGTTTC AGGT	539	874	Yes
299	deoxyguanosine kinase/deoxyadenosine kinase subunit	GTGCGTAACTCCTTTGAC GTGA	GATCGGAGTGGGCAAAAC GAG	GTTCGCCACGAAGTCGT AGTTC	520	757	Yes
302	glucosamine--fructose-6-phosphate aminotransferase, glmS ribozyme	ATGTCACAAGCCGGGGAG A	GTGGACCTTCCCTTTTCG CC	ACCGCTTCTTCGAGGTT GC	516	740	Yes
303	pgm phosphoglucomutase	CTGCGTACTGAATAGCCT AAAGGG	AGTCCTTGCTGACCAATA TCCCG	ACCTTTACCCCGTTCGC CA	289	586	No
325	malate dehydrogenase	CGTAAAACCATCCTCAGG AGGGA	TTCAAGGACGCCGACTAC G	CTGACCCATTCGCCGTC TTTG	557	822	Yes
349	ATP-dependent protease LA	GGTAAGACATCTTTTTGG GCTCCAG	GTGAAAAGGTCATCCTCA TCGTGTC	GTTCTTGTCGATTTCTT CCTTGACCTG	530	752	Yes
353	ribonuclease	GAGGACACCGACGAAGAC TTTG	GAGGATTACTACATCCCG CAGGAC	CAGACCGAACTTGACGA TGACG	450	767	Yes
378	TetR family transcriptional regulator	GTCCGCTTCATTCCTGCA CAG	CACGTTGTGCCTGTCGCT T	GCTGGTTGATTTCTGGG GCATAC	478	788	Yes
433	beta-lactamase	CTGCCTACGTGAGTGGGA AGTC	GCTCGTCACCGATTTTGC AGG	GAGAGCGAGGTGTTGCA TCTG	496	738	No

Table 2.4: *RT-PCR primers*

464	maltooligosyltrehalose trehalohydrolase	GCAGTTGCCGGAAGAAAGCGA	GCATCTACGAACTGGAACTGCCG	CGAAGTGGTTGTACACCACGTCC	462	756	No
515	hypothetical protein	CGCTCAACTGCTGAAACAATATTG	CGCGAGATGATAAGGTTCTAGCC	CGGCGCGGCTCATCTTTT	70	136	Yes
516	hypothetical protein	CAAGAAAACGGCCCCCACA	AAGATACCGCCCGAAACACC	TCCCACACCTGAGAGGTGTA			No
551	glutaryl-CoA dehydrogenase	AACACCCCTGACCTCAACGA	TACGACTACCTCGACACCCTG	ACCTCGTTCGTTTCCTCGTCG	542	740	Yes
555	glutamate-1-semialdehyde aminotransferase	GTCGGGTTTCATCGCTAAGTGT	CGTTACCTCGACTACATCGGCT	ATCACGGCGGCAATCTCGT	473	778	No
561	maltose ABC transporter periplasmic maltose-binding protein	TTTGCTGCCTTGACCTCCAC	ATCAGGCCAAGACCTTCGAGG	TGATGAAACTCGCCGCTTTGA	531	800	Yes
562	maltose ABC transporter permease	TGCCCCTGGGCCTCTTTT	CCTTTCCCTGGATTCGTGACTG	ATCACCGTGAAAAAGGCGAAGG	502	806	No
599	aminoglycoside N3-acetyltransferase	TAAAGCTTCTGACAGGCACCG	ACGTGCTGCTGGTGCATTC	CGAGGTGCAGGTTGGTGTT	459	697	No
606	groES co-chaperonin GroES	ACGGATTGAAGGCGCTGG	CGCCAAGGAAAAGAGCCAGC	TCGCTCAGCAGGCTGTAGTT	171	340	Yes
607	groEL chaperonin GroEL	GAGTCAGGTTCTGAGCCTGTTCG	AGGACAAGCTGGAGAACATCGG	GTCCTTGAGGTTGCTGATCTTCTTTTC	506	758	Yes
640	methionine adenosyltransferase	ACGGGCGACATCCCTCTC	GTGGACATCCAGAAAACCGTGC	GGAATGACTGCCGGAATCACATG	512	774	Yes
674	argininosuccinate synthase	TGACGCTTCCGCCTGACC	GTGTCAAGGCGTTGAACACCG	CCGTTCACGTACTCGATTTCGAC	535	770	No
678	argininosuccinate lyase	TGGGCTTCTACTCACTACGACC	CGAAAAGACCGCCGAACTCG	TCCAGCATCGCCACGTAAGC	507	742	Yes
788	branched-chain amino acid ABC transporter periplasmic amino acid-binding protein	GCAATCCGTTGACCGAATCTGGC	CCAAACTGGGCATGAATCTCTCGC	CTTGGTGATGACCGCCGTGAAG	455	696	Yes
813	1-pyrroline-5-carboxylate dehydrogenase	CAAACGCTGTCGCAAAGAACTCG	GAACTCCTTGGCAAGCACTACCC	GAAAATGGCGCAGGGGAAGTTCC	480	751	Yes
861	phytoene dehydrogenase	TGTTCTTTTTCCCTGATTTCTGATTTCCC	CCGCATTTCATCGAGGAACTGTT	AGGTCTTCTCGACGAAGTGAATCATC	535	809	Yes
868	purH bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase, pfl switch	CTTCTGGCATAGCAAAGACCGTG	GCTTTCCCGAGATGCTGGAC	GTTTTCGCCGTAACGGACCT	479	726	Yes

Table 2.4: *RT-PCR primers*, *cont.*

873	O-acetylhomoserine (thiol)-lyase	CCGTTCTGGACGAGGGATTTCT	CATCACGTCACGGTGCCAAG	TGAGCAATGTCACGAGCTGC	480	799	No
899	ribonuclease H	CCAGTTGTTTTGGATTCAATCCGACTT	AGCTCTACAGCGATGGTGCC	CCTTTGACCCACAGGAAGGTCA	355	647	Yes
932	polyprenyl synthase	GCATGACTGGCGTCTTTTCTCC	TGAGCGTGTGCGTGGAACTG	CAGGTCGTCTTGCATTTGAAAGG	437	692	No
951	sdhB succinate dehydrogenase iron-sulfur subunit	CTGCTCTCGCAGACGCC	CGTGCTCAACGAAATCAAGTGGTA	ATCAGCGAGGTGATCGGAATTTCA	567	799	Yes
952	succinate dehydrogenase flavoprotein subunit	GCTGAGGCGACATTCAGAGCA	GGCACATGTTCGATACCATCAAGG	TGGAACTGGTAGAACTCCATGTCC	523	760	Yes
954	succinate dehydrogenase, cytochrome subunit	CACCCATCACGAGACATGGC	GTTGTGTACCACGCCTTCAACG	CTTGATCCAGGCACGGTCAG	476	748	Yes
997	CRP/FNR family transcriptional regulator	CATCAAGGCGGACAGATTCGC	TATGTCAAGCCCGGCGAATACT	AGCAAAAACCCCCACCGC	501	732	Yes
1019	glycerol-3-phosphate dehydrogenase	TGTTCCTGAATGTTCGTCATCTGC	TGCTCGAAGGTCACGACTACG	TTCCTGCCCACTTTCCTCGT	506	759	No
1031	(S)-2-hydroxy-acid oxidase	GCTCTTTGTCCTCGCCTCTCT	AGCTTGGGCAGCCTGATGA	ACGCCGCCGTCGAGATAAAT	557	893	No
1055	aspC aspartyl-tRNA synthetase	CGGCACTGAACAACGACAACG	GTGCAATTTTTGGTGCTGCGC	CTCGAACACGCCGACCATAAT	483	691	Yes
1081	TetR family transcriptional regulator	ATCTAAAGTCAGGCACGTTTCCC	CTTTTGTGAGCCGCACCAACC	GTGCATCAGCAGCGAAAAGC	479	700	
1084	methylmalonyl-CoA mutase	CCTGATGAGGACATGCGAACG	GCTTTTCGACCGCCGAGGAAT	CGGAGATGGAAATCGAGTTGAAGC	453	770	Yes
1093	tryptophanyl-tRNA synthetase II	CGCTTAGCCGGGTGATTTCT	AGCTGTTCGTGCTGCTTGC	GCCGAGCGACTTGCTCATCTT	500	739	No
1105	DNA repair protein RadA	AAAAGATGGCTCAGCTTTTTGGC	GCAATTCCTGCGGCTACCAG	CGTCTGGATGGAGTCCACGATG	512	787	No
1120	butyrate kinase	GTCTCGACCACAAACTTCGCC	CACTTGCAGGTGAGCCTGAC	AACTGCTTGCCGACCTCGTAG	446	664	No
1122	prephenate dehydrogenase	GGCAAGCTTGACCGGACA	CAGCGTCAAGTCGGGCAT	CGGTTTTCGACCACCATGTCAC	414	794	No
1123	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	TAGCGTTCCACTCCTCACGC	ATCACGCTGCACGGCATTC	GCAGTTCCTTCACGCCATGA	553	811	Yes
1147	chorismate mutase/prephenate dehydratase	CCTCCACGAATAGACCTCTCCC	GCTACCCCACCTTTCACGAG	CCCAGGGTTTCCAGCAGAAAAC	535	736	Yes

Table 2.4: *RT-PCR primers*, *cont.*

1156	transcriptional regulator	AGTGGATGCCTTTCTCCCTTCC	TGACGCAGGTCGCCAAGAA	AGCGTGGTAATCGAGTGGGG	439	678	No
1160	uricase	CCTCAGTAGGTAGACATCTAAGTATCTGG	GTATGGCGATTTCGGTGCTGC	GGTGAAGGTGTGCTGAATCTGC	526	748	Yes
1225	mannosyltransferase	CAGATGCTTCAGGTTTCCTCTTTCC	CTGATTCACACCCATACGCCG	GCGGTAGTACGCCCCGATTT	549	845	No
1247	sucC succinyl-CoA synthetase subunit beta	GGACAAATGAACATCTGCGGCTC	CCAAGAACATCCTGGGCATGGA	CCGTAATCGCTGGCTTCGACTTC	511	792	Yes
1276	seryl-tRNA synthetase	AACCCCCAAGCTCCCTGAA	CTGATTCAAAAGGGCAAGGACCTG	CGTAGAGGCTATTGATCGGCAC	505	750	Yes
1279	Mn family superoxide dismutase	TCAGACGGCAATCCTTTCACTTCT	CCAACCACAGCATGTTCTGGC	CTTCGAGACTTCGTCCCAGTTCA	380	666	Yes
1304	spermidine/putrescine ABC transporter permease	TGCTCGTTGTCTCACGGTCG	CGAAGTGGTACTCGGTGCTGT	TCGTCGAGGCTCAGGGTAAAC	493	740	Yes
1376	hypoxanthine-guanine phosphoribosyltransferase	GACCGATTCACTGGGCTGACT	CACCTGATCTGCGTGCTCAAC	GCAGGTAACTCATGGTGATGCC	226	409	Yes
1384	TetR family transcriptional regulator	CATTCGTGAACTTTGGGACCTGAC	CAACGTCTACACCTATTTTCCCTCCA	CGACTTCGAGGAGCTGCTGAAT	386	673	Yes
1433	methionyl-tRNA synthetase	GGGGTTGCCCTTTTTCGTGG	GACGTGACGTTTGTGATGGGC	CCGATGTCCTCTTTCAGCATCTC	491	756	Yes
1456	ribose-phosphate pyrophosphokinase	AACATCAGCAGCAGCGGG	GAGAAGTTCACCAACGACAACCTG	GTCCACGATAAACACGGTCTTGC	564	843	Yes
1482	2-isopropylmalate synthase	CCTGCTTCTTCTTAGCCTTCCTCC	GTGATTTGGAAGGCGTCTCGC	ATGCCGAGGTCGTCGTGACA	463	755	No
1487	enoyl-CoA hydratase/3,2-trans-enoyl-CoA isomerase/3-hydroxyacyl-CoA dehydrogenase	GCACGTTCAATCTGTACGTCCA	TGCCGACGACAGTGTCAAGG	CGTCCCAGCCTTCCTGAAAG	623	844	No
1516	acetolactate synthase large subunit	ATTGTACTTCTTAGCCACGGGGG	GCTGACCTTTTACCCGGAAGTGC	TGGCATTTCTCAGCAGTTCTCTGG	497	746	No
1532	transcription-repair coupling factor	CTCCTTGTTTTGTACTCGCTTCGC	AACCTCGCCAAATTGCTGGAC	CAGCGTAAAGCTCTGAATCTTCTCG	555	785	
1573	pyrG CTP synthase	GTTCGACCGTCCAACATTCAGC	CCGTCAAAATCGACCCCTACATCA	TGATGCCGTAGCTCCTGAGC	504	754	Yes
1613	ArsR family transcriptional regulator	GAATATTTGGGCCACTCGCTGC	GGAAAAGACGGTCACGCAGTTG	CGCTGAAGCTGGTTTTCGTAGAC	198	426	Yes

Table 2.4: *RT-PCR primers*, *cont.*

1617	xanthine phosphoribosyltransferase	TGGACACCCAGTACAAGCCC	GACCAAGGTGCTGACCATCGA	GCCTTCTTCCTCACTCATCCGC	406	659	Yes
1662	tdh L-threonine 3-dehydrogenase	TCGTGACTTCTCCCAGCTCG	GAATACGTGGGTGTGGTCGC	CTCGTGCATGTCGAGTTCCTG	507	764	No
1698	tRNA-dihydrouridine synthase A	GTGCTGTTTTGCTGCCGC	TACGACGAGGTGAACCTCAACT	GGCATCTTCTCCAAACACGTCAC	495	800	No
1718	glutamate dehydrogenase	CATGACCAGCACCCAGCTC	CCTTCGAGCAGACCTTCAAAAATGC	CGTAGGTAAAGCCGTCCTCGT	501	813	Yes
1766	anthranilate synthase component II	CGGGTGACGTTGCGTCTTT	CTTTACGCTCGACGATGTGCG	TCGTCCTGACTGAGCACCTC	569	762	No
1778	3-isopropylmalate dehydratase large subunit	TCAGCGTGACCCGTGAGAAAC	TGTGCCACGAAATCACCACGC	TGTCGAGCTGGTCGATGTAGTCG	525	775	Yes
1817	phospho-2-dehydro-3-deoxyheptonate aldolase	GGCGAGTTCGGCTTTTATTGGT	GACCGCCTACTGGTGGTTGT	GCCCATCTTCGTCAATGGTGAA	496	807	No
1906	L-lactate permease	CCCCCTTTTTGGACCTGATCCA	CGATTGCGTGGATTATTGTCGCT	GTAGACCATGAAAAACGGCACGAAG	401	780	Yes
1907	fumarate reductase-like protein	TCGGGAGTGAGGCACACTC	GACTCCGAAGTTGTGGTCACG	CCTTCCGTACTCGCCAGAATCTC	527	799	No
2008	aspartate-semialdehyde dehydrogenase	GACATACGCGGGCGGTTT	GGGTCCATCAGCAAGGAAAAGG	GGGTTCGCAGGGTGGGAATA	502	753	No
2087	translation initiation factor IF-3	AGGTCCTGACCGCGCATT	TGCCGCCTCCTCGATTATGG	ATGTCCGTGTCACCAGCGT	401	625	Yes
2174	leucyl-tRNA synthetase	CATGATCCGCTCTGCGCTT	GAAAGCGGCCTGTACAAATTCGA	CGTTCACGACCTGCTCGTTC	463	765	Yes
2206	citrate lyase subunit beta	CCTACCTGCTCGTTTTGCTGC	GTGCTGCCGAAAGTGGAAGAGG	ACGTGCATCTGCTCCACCAAG	506	848	No
2208	lactoylglutathione lyase	ACCCGCTCCATTCTCTGCT	GACGGTTCCATGCTGCTGATTT	GGGATCGTCGAAATACAGGCTCTC	210	454	Yes
2217	terE tellurium resistance protein TerE	CAGCTGCCGACTTGGATTGG	GGACAACTCGCGCTGAGTCT	CCTGGAGTCGTCGAACACCA	425	792	No
2221	tellurium resistance protein TerD	TCGTTGGGAACCTCAGTGGG	GCAGCGTCTTTTTGCTCGATG	GACGGCCTTGAACTTCCAGTC	389	598	Yes
2225	tellurium resistance protein TerD	GCCCCTCTCCCTTCCCATAAA	CAACCAGCTTCGTTCGGTGG	GGAATTTCCACTCGCCGCC	320	542	Yes
2259	transcriptional regulator	GAAAGGCGGCTGCTGTTTTCT	GATCAACCCGCAATCGGTGG	TCTCCTTGTCCGTAGACCTGAAG	159	298	Yes
2285	A/G-specific adenine glycosylase	TGGTTTTTGCGTTCCAGGCG	GAGATTTTGCTGCAACAGACGCA	CTTTTCCAGCACGGCGTACTC	612	828	No

Table 2.4: *RT-PCR primers*, *cont.*

2303	chloramphenicol acetyltransferase	TGTGTGGAGAAAGAACATCGTTCTG	TGATTCATGCCGTCTGCCG	CAAGCTCCTTTCCCAGCACC	459	759	No
2306	MerR family transcriptional regulator	TTTACTTATCCCTCAGCGGACTCC	CGCTGCGGACCCTAAAGTACTAT	CGTCCTTCAGTTCCTTTTCGGC	300	560	No
2354	pheS phenylalanyl-tRNA synthetase subunit alpha	TTTTGGCGGGGCGTTTTTTTGA	GCACTCGACGCTAAACTTGCCA	GGGGTAATAGCTGGGCTGAAAGC	507	789	Yes
2361	acyl-CoA dehydrogenase	CGTTGTCAGGCAGGGAAATGGG	GCTTTGTCGAAACCGAAATCCAGC	GGCACGAGGAACAGGCTGAT	559	841	Yes
2396	TetR family transcriptional regulator	GCTTCCAGGCGGAAACAAACC	CGAAGCGATTCTCGACATTCTGC	TATCAGGCACCCGTTCCTTGA	392	636	Yes
2428	bifunctional nicotinamide mononucleotide adenylyltransferase/ADP-ribose pyrophosphatase	GGCTGGCGTGTGTGAAATGA	GTCTACATCGGACGTTTCGAGC	GAAAGCCCTCCAGAAAGGCATG	490	779	Yes
2438	endonuclease III	ATGAGCGACGACCAGAAGAAGG	TGATTTCGACCATCCTCTCGCA	TTGATGTGCAGCTCGTACAGATACG	466	799	Yes
2477	3-hydroxyacyl-CoA dehydrogenase	CAGAGCGTCAGGACAAGGAAGA	GACCGTAATTTCCTGGCGAAGG	GGGTCAAGCCGTACTTTTCCATG	526	762	No
2496	UDP-N-acetylmuramoylalanine--D-glutamiate ligase	GTGCATTTGTCCCAGAGCATTTGA	GGAAGACGAGGCGCTGATG	CCGCCGTGATGTTGAGTTTGG	473	741	No
2547	glutamyl-tRNA reductase	TCCTGTTCTCTGCTTCCCTCA	AACCGCACCGAGGTCTATCT	TTGACCACCAGCACGTCCT	464	735	No
2568	argS arginyl-tRNA synthetase	GGACAGTTTTGGCCTTCAGCACA	CGTTCCTGAACTTTTTCCTCGACGC	GCGAAAGCAGGTCTGAAGCTGC	461	791	No
2588	iron ABC transporter periplasmic substrate-binding protein	CACTGTCGCGCAACGGTTT	CGCAAGGTGACGATCAAGGC	CACGCCCATGAACGAGTTGG	489	784	Yes
2630	thyA thymidylate synthase	CATGAACCTAGAACCCTACCAGCA	ACAGTACCTCGATTTTCTCCGCC	TAGAACTGAAAGAGCAGGTGGCAG	453	785	Yes
2635	pyruvate kinase	TGCGGATGGCTGCCTCTTT	AAGGGCGTGACCATCGGTA	GCCTGCGGCTTCTCGATTTTG	500	741	Yes
2637	eno phosphopyruvate hydratase	CCTTATAGTGTCGGGGCTACGC	TACATGGGCAAGGGCGTG	TGCTGAATGGCTTCAAGCAGC	509	766	Yes

Table 2.4: *RT-PCR primers*, *cont.*

Gene #	Gene Name	Verified by RT PCR	Primer Sequence
	Random Dodecamer		NNNNNNNNNNN
98	(100 is SSB)	N/A	GCCCGAAGCCTTGATGGTGTA
105	LEA-76 Family protein (Hypothetical)	N/A	CATTGTGAACGTCCGCTTTGACA
129	dnaK	Yes	CGCAGGTCGAAGTTGTGTTCC
139	GTP-binding protein HflX	N/A	CAGACGCAGGTCGGGAAACT
153	riboflavin-specific deaminase/FMN switch	Yes	ATTCGGGCTGGTTCGGCCCAGG
229	endoglucanase	N/A	AAGGACGCTGAGGTGTCCCTTAT
302	glucosamine--fructose-6-phosphate aminotransferase/glmS switch	Yes	ACCGCTTCTTCGAGGTTGC
309	Elongation factor Tu	N/A	GTCGTAGGCCAGGGTTTCGAT
326	Hypothetical Protein	N/A	GCTTCGCTGCCTTCGAGTTC
349	ATP-dependent protease LA	Yes	GTTCTTGTCGATTTCTTCCTTGACCTG
378	TetR family transcriptional regulator	Yes	GTGCATCAGCAGCGAAAAGC
423	DNA damage response protein A, DdrA	N/A	CTTGGGCAGGTCGTAGAGGTAA
435	Cytochrome complex iron-sulfur	N/A	TTCTGGCCGTCTTCCACGTT
561	maltose ABC transporter periplasmic maltose-binding protein	Yes	TGATGAAACTCGCCGCTTTGA
640	Methionine adenosyltransferase	Yes	GGAATGACTGCCGGAATCACATG
694	Hypothetical Protein	N/A	ACCTGCTGCATGGGACTAATCA
868	purH bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase, Pfl switch	Yes	GTTTTCGCCGTAACGGACCT
899	Ribonuclease H	Yes	CCTTTGACCCACAGGAAGGTCA
950	NADH Dehydrogenase	N/A	TGCGGTTGACGAAGTTGAAGATGT
971	electron transfer flavoprotein subunit beta	N/A	CGTCGATCTTCTCATCGGTTTCCA
1126	recJ	N/A	GACGACCAAACATTCGGGAAAGTT
1198	GTP-binding elongation factor family protein TypA/BipA	N/A	GCGTCGTTGCGGTCGATCTT
1247	sucC succinyl-CoA synthetase subunit beta	Yes	CCGTAATCGCTGGCTTCGACTTC
1279	Mn family superoxide dismutase	Yes	CTTCGAGACTTCGTCCCAGTTCA
1310	RecX	N/A	CTCGACCAGTTCGGGGTCTT
1356	Transporter cluster	N/A	CGAGCAGGTTGAAGTGCTGAAAGA
1573	pyrG CTP synthase	Yes	TGATGCCGTAGCTCCTGAGC
1617	xanthine phosphoribosyltransferase	Yes	GCCTTCTTCCTCACTCATCCGC
1663	Hypothetical Protein	N/A	CCTTCTTCCTTGGCGGTGTTGAA
1768	Hypothetical Protein	N/A	TCATGCCCTTGTCCACGTACAT
1771	uvrA	N/A	GTCGTGAATCTCGGTGACGGTA

Table 2.5: *5' RACE primers*

1778	3-isopropylmalate dehydratase large subunit	Yes	TGTCGAGCTGGTCGATGTAGTCG
1816	Hypothetical	N/A	GCGGTACGACCCACCATAAGA
1817	phospho-2-dehydro-3-deoxyheptonate aldolase	No	GCCCATCTTCGTCAATGGTGAA
1857	organic hydroperoxide resistance protein	N/A	AGTCGCACGTCCACGTTGTT
1913	DNA gyrase	N/A	CGTGTTTGACGTTGCTGGCAA
1940	Hypothetical heat shock protein	N/A	GGTCGTCAGGTTGGCTGGTA
1987	Hypothetical Protein	N/A	GGCCGTCATGCCGAGGATTA
2075	Ferredoxin	N/A	CAGGCCCTTGTCCTGCAACTT
2208	lactoylglutathione lyase	Yes	GGGATCGTCGAAATACAGGCTCTC
2221	tellurium resistance protein TerD	Yes	GACGGCCTTGAACTTCCAGTC
2275	2274 - Hypothetical protein (2275 is uvrB)	N/A	CCTGTAACCAGCACACCGCTT
2361	acyl-CoA dehydrogenase	Yes	GGCACGAGGAACAGGCTGAT
2438	endonuclease III	Yes	TTGATGTGCAGCTCGTACAGATACG
2569	Hypothetical	N/A	CGGACTGCCCCAGGTGTTTA
2577	S-Layer Protein	N/A	GCCAGTTCCTGAATGGCGTTTT
2588	iron ABC transporter periplasmic substrate-binding protein	Yes	CACGCCCATGAACGAGTTGG
2630	thyA thymidylate synthase	Yes	TAGAACTGAAAGAGCAGGTGGCAG
2635	pyruvate kinase	Yes	GCCTGCGGCTTCTCGATTTTG
A0202	A0205 - Sensor histidine kinase (may be on same mRNA as A0202 Cu/Zn family superoxide dismutase)	N/A	CCTGAGGTGCTCGGCAAAGTT
A0301	A0303 - Hypothetical protein (may be on same mRNA as A0301, methylamine utilization protein)	N/A	GCAAAAGCGTTGCTCAGGCAAAA

Table 2.5: *5' RACE primers, cont.*

Gene #	Gene Name	Sequence (UTR+60NT Coding)	*Genscript ID*
98	Ribosomal protein (100 is SSB on same mRNA) ((99 is a DDR gene with a RDRM that DdrO binds…))	GTTGCGCCGTCCGGCGAGGTTTGACTTTGCCGCGTGACCGCACCGACAGCAACACTATGAGGAGGTGAATCA TGAACCAGTACGATCTCAACCTGATCCTGAACCCGAACATCAGCGCCGAACAGGTGCAG	*53*
105	LEA-76 Family Protein (Hypothetical)	ATGAGCGACAAAAAGAATATGTTCGAGAACATCGCCGACGCCGCCAAAGCCGGGGGCCGACAGCCTCGCGGA CCGGGCCCGCGCCGTGGGCCACGACATCGCTTCGCACGTGGGTGATGAGGCGGGCAATGCCCAGGACCGTGC CGAGGCCGCTGCCGCCCGTGCCCGC	*51*
129	dnaK	AGTTCAAGACTGCTGCCCTAAGGGGCTGCAAATCCCTCTTTCCCAAGGAGTCAACACATGGCTAAAGCTGTCG GAATCGATTTGGGTACCACCAACTCCGTGATCGCTGTGATGGAA	*17*
139	GTP-binding protein HflX	AATGCCTGAACGTGTCTTTTTACCGCCCGCACGATGTCGCCGGGCTGTAGCTGCCGCTCCTGTGTCCATGCTCG TGGACCGGGAGACCCGCAGGCAACACCGTCACCCACAGAACTGACTGGCCCCCCGCGTGAGTCGAGCCGACG CTCCTCCGCTTTGCCCCCTCCTCTCACGGCCCACCTCGTGGCCCCAGAAAAGGAAGTGACCCAACATAGATAA AGTTCACGGTAACCTCTCCGGCCTGCGCCCGGCCCAGAAAAAAGCCCTGGAAAACTTATATCGCCGCCGCAT CGAGCCGGGGCGCATCGGTTCGCCCGAACTGGCCCGTAACCTGGCCGAACTGTCGCATGACGTGCGGCGCGA GGTCGGCGTGCTGATCGACCGCCGGGGCCGCGTGATTTCGGTCAGCGTG	*45*
153	riboflavin-specific deaminase/FMN switch	CTTCAGGAACCGACCTCTTTCGGGGCGGGGCGAAATTCCCCACCGGCGGTAAGTTCTCCCGAACAAGCCCGC GAAGCCCGCGCAAACCGCACCACGCGCCGGGCCCGATGCCGCGCAACTCGGCAGCCGACGGTCACAGTCCGG ACGAAAGAAGGAGGAGACGCTCAGCTTGCCCCCCAGCAGGCGGCGTCCGCGTATGAATACGGAACATCCGCC TGACGCGGAATACATGCAGTTGGCCCTGAACGAGGCCGCC	*69*
302	glucosamine-fructose-6-phosphate aminotransferase/glmS switch	CCGAACCCCACACGAGTGGGCGGGGTTCGACTTCTTCGGGCAGCGCAAGGCCCCGGCGACACGTGATGTCAC AAGCCGGGGAGACGAGGTGGAGGTCAGCGACTTTTCTGCGGATGCCTCCAGGCCCCGGTGAACGGGCCTACC CGGCGCGTGCTTTGCCGCTCTGAGTCAAAGACTCCGGCAGGCAGAACCACGCGCAAGCCCGGCGATAAGCCC CGCAGCAATGCGGGCATAAGGCCGGGCAGCTCACCACACCCCAGCAAGTGGCTGTGCGCGGCGCCGTCCCCC GGTGGACCTTCCCTTTTCGCCCCGGTGCGGAAGCG	*27*
309	Elongation factor Tu	CGGGATCTTCGACGAACAGGTGACCGCTTACCTGAGCCGCTGAGCGGGTTTTCGGGCAAGAGTTAGTTCGGG GCCAGGAACCTATACGGAACCTATACGGGAACTGGCCCTTTTTGCTGGATTCCCGGCCCGGCGCTCTTGCCCT TCTAGGCCTTTTCATGTAAGCTAAGCAGTCGGTGCGCCGTGGCCTCTCAGCCCGGTGCAACGTGCCAGAGAGC AGGCGGGACACCGCCCTGGCGAGAATCAACCCAATGTGGGTCACACCACCAGAACCTCGCGGCATGCGCGGG GGAAGCCCGATCAAAGGGACGTTTTTACGTGTGAATCCACCGCTTGGAGGGAGTAAGACAATGGCAAAAGGA ACGTTCGAGCGCACCAAGCCCCACGTGAACATCGGCACCATCGGTCAC	*77*
349	ATP-dependent protease LA	CAGGTTTTTTGGTAAGACATCTTTTTGGGCTCCAGGCCCCCTTGCCAAGCCACCCCTTCGCGACGGCCCCCTGC GGGTCAAAAGGAGCGTCAACGGATGCCCGACAGCACTGCCCTTCCGACCACCATCCCCGTCTGCCCGGTTCG CGGCAGCGTC	*19*

Table 2.6: *Final Candidate 5'UTRs*

378	TetR family transcriptional regulator	GTCCGCTTCATTCCTGCACAGCCGTTCATTCCTGCACAGTCGGGAAAGTGCCGCCGGGAGAGTGCCGCCTGTG CATCCACACCGCGTGACCCGTAGTTTTTCCTCCTCGCATCCGCTCGGATTGAAGCTGAAACTGCCAGATCCAA TCGGAATCCGTATCAGCGCCGGACGTTCGGGCTGGGGGGGGCGCTGGGAGCGCAGTGGCCGGAGCCGCCCCC TCATGACTGACCTCTGCTAACTTTTACTGACCTGAGGTTCCCTCAGCAGGTGAATGGTTTCACGTTGAGAAGA CGAACAGCCGTGTAGGGTGCACGTTGTGCCTGTCGCTTCCTCCCGTCCCCTCCGTGCACGCAGCGCACTCGAA AAG	*13*
561	maltose ABC transporter periplasmic maltose-binding protein	GAGCCATACCGACTCACCACCGCCGGGCCTGTCCCGCCCCTTTGCTGCCTTGACCTCCACGGAGCGGCCTCCC CCTTGCTTGGGCCGTGCTCGCTCACGGGCCGGGGCCGCGCGGTGACCAACCCCTTAGAGCTCTCCCTTTTCGA GCCCTTTTCAGTGGAACCGGTTCCCTCTTGTGTGGGAGCCGCCCGAGGAGACCCATGAAAAAAGCACTGACC CTTCTTTCTCTTGCGTTGCTCGGCCAGGCCAGCGCCGCGACC	*23*
640	Methionine adenosyltransferase	GCTGCCCGCTGTGGCACGCGGCGACGGGGAACCGACGGGCGACATCCCTCTCCAGGCCGTTCTGCGAGACGC TGCCCCGCGCCCAAGGAGAATCTATGAGCAAGTTTGCTGTGCGTAAGTTCTACACCTCGGAATCGGTGTCCGA AGGGCACCCC	*7*
694	Hypothetical protein	GAGTCTGACGCAATCATCACAAGGTCGGGCCATCAAGATTCACCGCTCTGCCCGCGCAACCTCATCCCCGGAT GAGGCTTTCGTCTGGAGGGCGGTTGAAGAGTGAAGGGAAAGGAGGAGGGAGTTGGACGTTTCAAGTCGAGT CTTGAGTGAACTGGCGAGCCGCGAAGCCGCCCTCGATCAG	*35*
868	purH bifunctional phosphoribosylaminoimidazolecarboxamide formltransferase/IMP cyclohydrolase, Pfl switch	GTCTGCGGAAGGTGACTGGCGCGAAAACGTGGGAAGACCACGGGGAAGCCTTCTGGCATAGCAAAGACCGT GCGCCTGGGTCGGACGCTTCCTGCATGTGCGGGGGCGTTTTCCTGTGTGGAAAACGGGAGGCAACATGACGA AACGGGCGCTTATT	*3*
899	Ribonuclease H	CTGCTACGGCTTCCGATTGAATCCAGTTGTTTTGGATTCAATCCGACTTGTAAAGCTGCGCAGCAGCGCGGAT GCGAGTAGAAAAAAATACGGATTTCGCGGTATGGAAGCGCAGGCGGTGCTTTTCCGACTGTGCTGGAATGGA GCGCAGTCCGTATAAGCAGTTGACAGCCCCCGCCCGCGCTCCCTACAGTTCGCGCCTATGACTCGCCCAGGCC GCCCCTCTGCCCGCAAAAAGCCCGACACCTCGCGTGACCTGCTG	*67*
950	NADH Dehydrogenase	GAACGCCGGGCGAAAGTCCCGGCACCTTTTGCGGGCTGCTTTTTTTCGGCCTCGGCTCGTGAATACTTTCACAG GAGTTTTGAAGATGAAAACGCTGATTCTTGGTGCGGGCTACGCTGGCCTTGCCACGACCACTTCCCTCAAG	*47*
971	Electron transfer flavoprotein subunit beta	GTGCTGGAAGTGGTGGAAAATCTGAGCCTCACACGCTAACTTGTACCGAGTACAAACTTGTGTGGCGAGTCC GGCAAGCCGGTAAGCAAGTCAACATTCAGCTGCCCCGACCGGCAGCATGGTGAAGTGGAGGGAGACGCATG AATATTCTGACTCTGGTACGGCAGGTGCCCGACGCCGAAGCCCGCGTGAAAGTGGCC	*57*
1126	recJ	CCTCGGCCCCGAGTGCCACGGAGACGGTGGCTGTAGGGGGGAACGAATGAGCCGGCCTGCCCACTGGCTGCT GGCGCCGCCCGCGAGTCGGGACGCTCTGCTGGCGACCATGCGCGAATGGCAGGTGTCGCCCCCAGTGGCGCA GGTGCTGTGCGGACGTGACCTGCGGACGGAGCTGCTCGCTCTCCCCCTAGAACTGACGCCCAACCCGGCGCT GCGTGAGGCCGCCAGCACATCGTGGCAGCGGTGCGCGAGGGAAAGCGCATCCGTATTCACGGTGACTACGAC GCCGACGGG	*73*
1198	GTP-binding elongation factor family protein TypA/BipA	TTGGGTATCCCGCCCATCTTCCGAACTCTGGAATCGCGGCGGGTTTTTTTGTGCCAATTCAGGCTCATAAGGG CTGATGGCTGAAGGCTAAGGGCCGAAAGCTAAAAGCTCCGTACCGAGGGAGCAACTATGGAATACAGAAAC ATCGCCATCATCGCGCACGTCGACCACGGCAAGACCACGCTGGTC	*81*

Table 2.6: *Final Candidate 5'UTRs, cont.*

1247	sucC succinyl-CoA synthase subunit beta	ATCGGGACAAATGAACATCTGCGGCTCCCCACACCCACCGGGGGGTCATTCCCACCACAGGAGGCAACACCG TGAAACTCCACGAGTATCAGGGCAAGGAAGTGCTGCGCGACTTCGGCGTGAACGTGCAA	*21*
1279	Mn superoxide dismutase	CAGACGGCAATCCTTTCACTTCTCTCTTGCCGCCACAAGGAGCACTCATCATGGCTTACACTCTTCCCCAACTG CCCTACGCTTACGACGCGCTTGAGCCCCATATCGAC	*61*
1356	Transporter cluster	TCGGCCCTGGCACGTGCAGGGTCACCTTTATCCAGAGTCGGCGCAGGGACCTGGCCCCATGACCGCCGCAGC AACCGGCCCTCATCACGGCAGCGGTGCTTTCCAGACCCGCGCGAGCAGCGCCCGACGATGGGCGGCGCCGCG GGAACGATAAAGGAAGGCGGGTCCTCTTCGCGGGTTCCAACGGACGGCTCAGCCCTGGGCGTCCCCTTCCAG ACTTCTTTTCGTCCAGGAAGGGGACGCCCGTTTTGGGCCGACCTCTCCGCTCTCCCCACCGGAGGCCCGCCCC GTGACCTTACCGTCCTCCCCCCCAGCCTTGCACTTCGAAGGCGTCAGCAAAACCTACCCC	*43*
1573	pyrG CTP synthase	AAGTATGCATGGGACGGCATTTAGACTTTTAGACCGTTCGACCGTCCAACATTCAGCTCCTTCCCCGTGCTAT CCTCAAATCCCGGAGGCCACACGCCCCGGCTTTTTTTGTTTTGGCGTGTGGCACGTGCCTTTCACAGTCCCGAC CCTCGACCCTTAGACATTCCCGGAGGGAATCCATGAAATACATCTTCGTGACCGGCGGCGTGGTCAGCAGCCT CGGCAAAGGCGTGGCGAGT	*65*
1617	xanthine phosphoribosyltransferase	ACCGGGTGGACACCCAGTACAAGCCCTGGCAGGCGGTCTACGAGTACGGCCCCCGGACGAACTGAGCCGATC GAACTGAGCCGCCAGCGTCCATTACACTTTGCCGTATGCAGTCATTGGTCGAGGCGATTAAGCGCGAAGGCA GGGTTCTTCCGGGCGGGTTTCTC	*15*
1663	Hypothetical protein	ATGATCAAGTACCGTAAGCAAAACGCTCTGGAGCCCGAGAAAGACGCCGAGATCACCCTCAACTTGACCCCG CTGCTGTTTTTCACAGTCGGCTACGTCGCCATGAAGGCGCTGCTGGGCAGCGTGCGCCGCCTGAGCTGAGCGG CTCCCCTCCCCCGGCGGCCCCATTTCGTGCTGAAGCGGGCCGCTTCCTGCTGTCTGTCTCCTGACAGGCGGGT GCGCTAAGGTGTGCTTTATGACCAATCCTCCTCAGACTCCGCAGCAGCCGCAGCAGTTCAAGAGCGTG	*39*
1768	Hypothetical protein	GCTATGAAAAAGACACTGATGCTGACTGCTTTTGCGCTGGCTGGACTCGCTGCCGCGCAAACCGCTCCCGCGA CCACCGCTCCGATGACCAGCACGACCACGATGACCGGCGACGCGATGACGCCCGCCACCCTGTCGGCCAGCG ACAACTTCGCCCGCGCCCAGGAATATGCAGTG	*49*
1771	uvrA	GTTTGAGAACGTAGGGCCAAGGCTGCCTTACGCGCCTGCCGTAATCATGCCCTGTTGCCAAAAGGCGTGCAG GACAGCCAAAAGGAGCTTCATGCAAGACAAACTCATCGTGCGCGGCGCCCGGGAACACAACCTCAAGGACA TCACCGTG	*33*
1778	3-isopropylmalate dehydratase large subunit	ATGTTCAGCGTGACCCGTGAGAAACCGGCCCCGGCGGGCCTTCTTCTTCGTCTGGGGCGTTGAGCGCGAGTTT CCCAGCCGTTCTTCACGTGCCCCCCGCAGCCGAGTAGCTGAGGGGGCGTTTTTCGTCCGAAGGATTGATCTGG GGGAGGTCCATATGGGAATGACGATTGCCGAGAAGATGCTGGCCGCCCACAGCGGGCACGACACCGTGGTG	*29*
1817	phospho-2-dehydro-3-deoxyheptonate aldolase	ATGACGGTTCAGCGCGCGCCCCTGTGCCCCCACTGCGGCGAGTTCGGCTTTTATTGGTTCGGTTCCCACCGGG CCTAGCCTCGCGCTGTTTTTCCCCGCCCGGTGACGCCCCGCAAAGGCTCGCCGGGTTTTTCTTTTTTCCCCAGC TAAATAACCCAGATAAAAAAGGACGTGTTCCCATGACCCAGCCCCCCGCCCCCTCTACCACCGAAAACCTCA ACGTGACCGGCTTCACGCCG	*75*
1857	organic hydroperoxide resistance protein (OHRP)	GGGCCGCCCCAGGTTTGGATTCAGGTCCGTCTCACTTTCAACTGATGCGGGCGAGGGAATACTCCGGGTATGG CAAACGTGTATACCGCCGAAGCAACCGCCACGGGGGGCCGCGCCGGAACGACCCGC	*71*
1913	DNA gyrase	AGTTGAGGAGAGCCGTTTTATTACGTGATTAACATAATGTGGTAGACTGGTTCTATCCCCCGGCGAGTGCCGG GACCTCTCACCTTTTTCCCGGCTGCGGACTGAGCTTTTCCGCGCCGCCCCGATTGGAGTGCCATGACCGGAAT TCAACCTGTTGACATCACCAGCGAAGTCAAGACCAATTTCATCAACTAC	*37*

Table 2.6: *Final Candidate 5'UTR, cont.*

1987	Hypothetical protein	GGAGGCGTAACTGTGACACTGACACCTATGAAGCACATTCTGTTTCCGACCGTGTCTGCCGCCGACGCCTTTA TCGCCGACCTGCAAAGCCGTGGTGTGGTGCAGCCCCAGGTGGGCACCATGAACATGAGCCGCCGCGTGCAGC AGGCTGCCGGTGACACCATGAGCACCGGCACCACGACCGGCACCGTGACGACC	*79*
2075	Ferredoxin	ATGACCGAACCCCTGATTGCCCAGCAGCAAGGCCAGCCCGTCACCATTGCCGTGGAGGGCTACGGCGAGATT CAGGCCCACTCGGGCGAGCGGCTGGTCACGGCACTGGAG	*59*
2208	lactylglutathione lyase	CGGCCCCCGAGCCGCGCGTGGTGCTGCCCTAGAGCAGTTCTCCGAATTACGCGGGTGTCGGCAGCAGCACCG CCACCCGCTCCATTCTCTGCTTCGCAGCTTTGCAAGTCCGTTCTGCTCAGTGGTTTGCACTCGCTCCGCTCGCC AAAAAGCTGTGCCATCTTTTTGGCAAATGTTCTAACCTAAACTGCTCCCATGCGCGCCCTCGAGACCTGCCTTT ACGCCGACGACCTCAGCGCCGCCGAAGCCTTTTAT	*9*
2221	tellurium resistance protein TerD	CTCCGGCGGAGTTCGGCCTCTAAAGACACCCGCTTAAATACCCCCGCTGGGGGCCCTCGTTGGGAACCTCAGT GGGGACGACAGTCACATAGCTTCCTTTTATACTCAGCGCTGAATATCAATTCAGGAGGTTCAAAGATGGCAGT TTCTCTTTCCAAAGGCGGCAACGTTTCTCTTTCCAAGGAAGCCCCCGGGCTC	*5*
2274	Hypothetical protein (2275 is uvrB)	AAAGTCCTGTGGCCTTTCTTAACACGCCCTTCGGCTTACGCGCTGCTGCGCGATGTCTGGGACTTTCCGCCGCA GGCCGACTGGCAACTCGCGCAGCTCCGCACGGTGCCGGGGCTGCGGCTGGTGGTCCTGAAAAGTGACGCCGA CATAAACAACTTTCTGGAGCAAATCCAATGACCGAACAGGCCCAGTACCGCGACCCCCGCCTCGTCATCCTCT ACGACCGGCTCAAC	*83*
2361	acyl-CoA dehydrogenase	AGCAGAGCCGAAAAACGTTGTCAGGCAGGGAAATGGGGCGGGGGAGTCGGCCTCATGACATTCGCCCCGCA AAGTTTTATACTCGGTTCAATTTCAAAGCCCAGGCTCTCCTTCAGGGGAAGAGTCTGCCCTTACACGGAGGAA TCATGCCCAGCTACAAAGCCCCCACCCGCGACATGAAGTTCGTCATGCAGGAACTGCTCGGC	*25*
2438	endonuclease III	ATGAGCGACGACCAGAAGAAGGGCTACGACCCCGCCAACCAGTCCCCCGCCGAAGGGCAAAGCCACGCCAT TCCAGCGCAGGACCGGGGCAAGGACCCGAACATTGACCCCGCCGCCAAGGACCAGCCGGCGGAAGGTGGAC GCGAGGAAGCCGAAGACGGCGCCCAGCAGTCGTCGTGACCCTCTTCGGCGACGTGTCGGGAAAGGGGGCAC CGCTCAATGCCGCCCGCCCCGCT	*31*
2569	Hypothetical protein	AGTGCTCGCCCCCTCCCGTGACCAGGCCCCTGTGACTGACTTCCTTTCCTCTACCCCCCTGACGCCTTCCGGCG TGCGCTGGCCCGTGTGCTCGCGCTGCTGGTGGGTTTCGGTCCGCTGGCGGCGCACGCTCAGGCGGCGGGCACA GCGGCGGTGAAGCCCCTGGTCGCCACGGTCGTACGCCGTCGGCCTCGCGCCCAAACGCGGCGTGCGCTGGAC GCGGCTGGGGAACCCCAGGTCGCCCTCGCTGCTGCAAGTGCCCGCCGAT	*41*
2577	S-Layer Protein	GATCTGCCCTCTTCTGTAGCCACAGGCTGCGAAGGGTGGGTTGCGCGAAGAAGGCGGAGTAAGGCCCAACTT CTCCTGGGAGAGGAAACATGGCAACTCGCCCCATCAACATCCTGCGTGCTGGAAACGCTGCTCGACGCGCAC TGAGATTGGGGGTTCCCTATGAAGAAAAGTCTGATCGCTCTTACCACGGCGCTGTCGTTCGGCCTCGCTGCCG CCCAG	*55*
2588	Iron ABC transporter periplasmic substrate-binding protein	AGCAGGCGCAGGGAAAGTCCGGTTTCAGTCCGGCACTGTCGCGCAACGGTTTTCAGTCCGAACACCTCGCCT GCTCGCGCTGCCCCTCTCATTTTTGGCGGGCAGCGTGCGGCCTGACCTCTCGCGGAAAGAGGGACCGCCCGTG TCGCTTCTC	*63*
2630	thyA thymidylate synthase	ATGAACCTAGAACCCTACCAGCAGGCTTCGCCCGAGCGCCTGCGCCTTTACCTGTCCCTCTTCGCCATGGTCG CGCTGACCGTGCTGCTCGGCGGCCAGCGGACGGGCGACCTCGGCGTGCGCGGCACCCTACGGCGCAGCGCCT TTGCACGCGGCATGTTTGCCGACATCGGGGCCATCAGCACGCTGGGGGCGCTTTATCTCGTCCTGAACGGCAG AACCTCCGTGCGGGTGCCCGCTGCTATCGCCACGTTGTTTGTCGGAAGTTTTGCGCTGCTGCCCGCACTGGCCT ACGAAGACTGGGCCGCCATGCAGGAGCGCAGGCAGTGAAACAGTACCTCGATTTTCTCCGCCACATCCGCGA CCACGGGACCGACAAGATGGAC	*11*

Table 2.6: *Final Candidate 5'UTRs, cont.*

As a prelude to experimentation, structural conservation was analyzed for the 41 5'UTR candidates based on the rationale that structural conservation among closely related organisms to *D. radiodurans* of the 41 5'UTR candidates could reveal structures or motifs critical for regulation of gene expression. An RNA structural conservation analysis was conducted using LocARNA, a published algorithm that uses an alignment tool to determine sequence conservation and a covariance model to determine structure conservation (47–49). For the analysis, the genomes of 45 organisms were examined, which included all members of the *Deinococcus-Thermus* phylum as well as some increasingly distant organisms, Figure 2.3. Each candidate 5'UTR was analyzed using the NCBI BLAST alignment tool against this specific list of organisms to collect a set of sequences with apparent sequence homology. While four of the candidates (DR2221, DR98, DR971, DR1857) did not have any sequence homology with the selected list of organisms, most candidates contained structurally conserved motifs. Many of these structured motifs appeared quite complex, displaying distinct base pairing and hairpin structures. Results from this analysis indicated the possibility that some of the experimentally confirmed 5'UTR regions could be regulatory in gene expression.

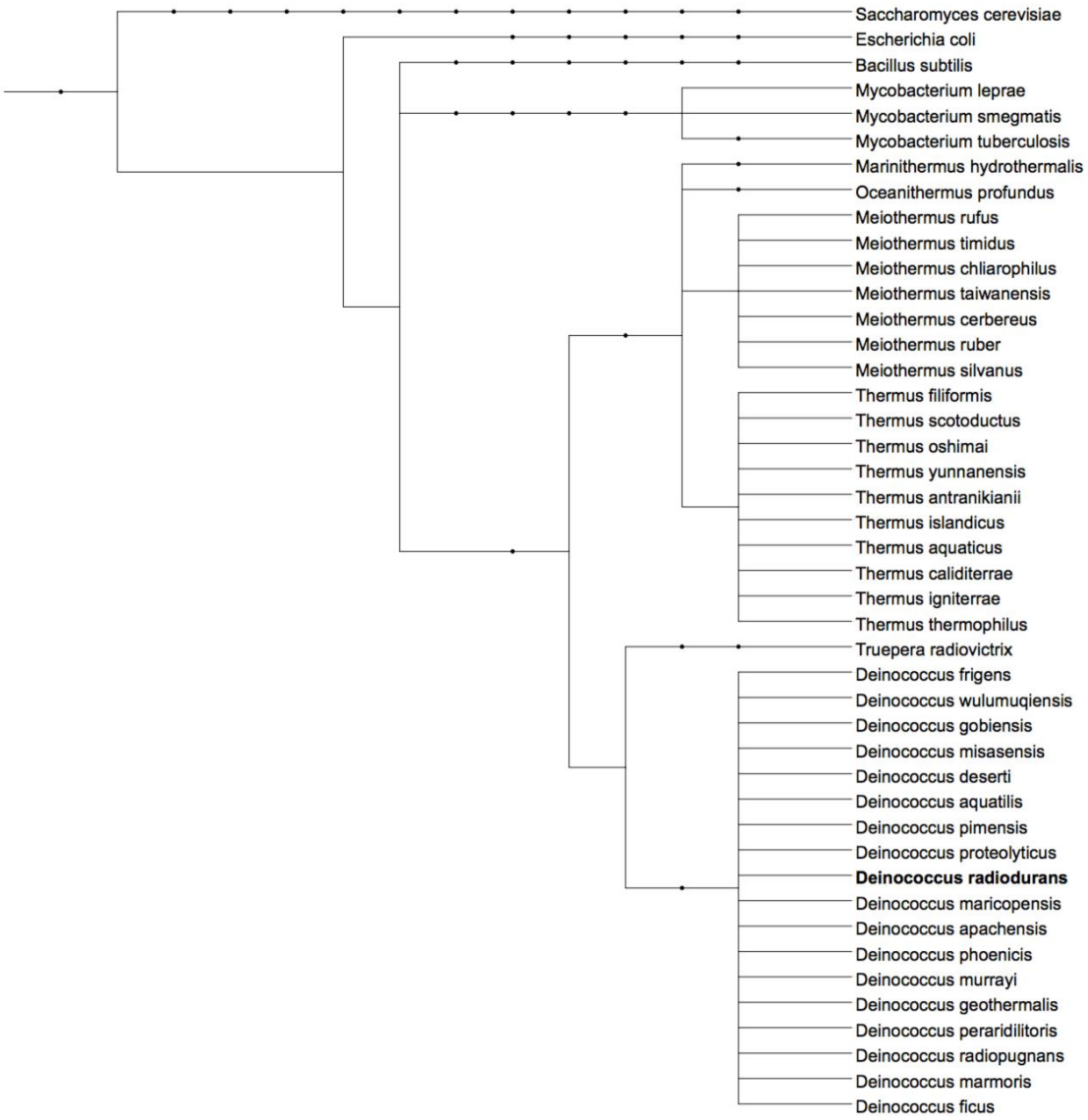

Figure 2.3: *Phylogenetic map of organisms used for conservation analysis*. List of organisms indicating evolutionary distance for organisms used in conservation analysis with LocARNA.

This bioinformatics approach revealed the presence of three widely conserved bacterial riboswitches that have been deposited in the RFAM database (the FMN switch (DR153), glmS ribozyme (DR302) and a putative pfl riboswitch class (DR868)). These riboswitches, contained in the 5' UTR, have been demonstrated to control regulation of their adjacent genes in response to small-molecule ligands such as flavin mononucleotide (FMN switch), glucosamine-6-phosphate (glmS ribozyme) and ZMP (pfl riboswich class) (36, 61, 62). These results served to validate that our method was capturing true regulatory 5'UTRs (glmS and FMN have been experimentally confirmed in *D. radiodurans R1* (36, 61) and provided initial experimental proof of the presence of the pfl riboswitch in *D.radiodurans*. While these riboswitches have known ligands, possible connections to a response to radiation have not yet been elucidated. As for the remaining 38 5'UTR candidates selected for further study, many are associated with differentially expressed proteins during ionizing irradiation and oxidative stress**.**

Design and validation of a fluorescence screen for detection of stress-responsive regulatory 5'UTRs. To test the independent ability of the candidate 5'UTRs to regulate gene expression in *D. radiodurans* under irradiation, an *in vivo* fluorescence screen was designed in the context of the pRadGro shuttle plasmid. This vector has been shown to be successful in the overexpression of exogenous proteins and is able to replicate in both *E.coli* (where it confers ampicillin resistance) and *D. radiodurans* (chloramphenicol resistance) (45). This pRadGro-GFP reporter plasmid contained an insertion site for a 5'UTR element, a constitutive *D. radiodurans* groES promoter, and a codon-optimized GFP gene (as presented in methodology)

Initially, we established this screen in the context of a well-characterized synthetic theophylline riboswitch (50). This riboswitch is present at the 5'UTR and activates

expression of the downstream gene when bound to the small molecule, theophylline, shifting the structural conformation to expose an RBS and initiate translation. Fluorescence was measured using a flow cytometer and gating for the correct population (Figure 2.4, A). A consistent fluorescence increase (~4 fold) was observed following induction of the pRadGro-ThRS-GFP construct with 2 mM theophylline (relative to a DMSO-only control), Figure 2.4. Although modest, this detectable shift is notable given that *D. radiodurans R1* only keeps 4-10 copies of the plasmid and the synthetic theophylline riboswitch was not optimized for activity in this organism (63). We therefore reasoned that this assay could readily detect shifts in translational activity mediated by a 5'UTR region post-stress. During writing of this publication, another lab reported the development of a similar *in situ* fluorescence assay to evaluate promoter activity in *D. radiodurans* (64). However, it is an important distinction that their method aims to determine promoter activity, while ours interrogates 5'UTR regulation.

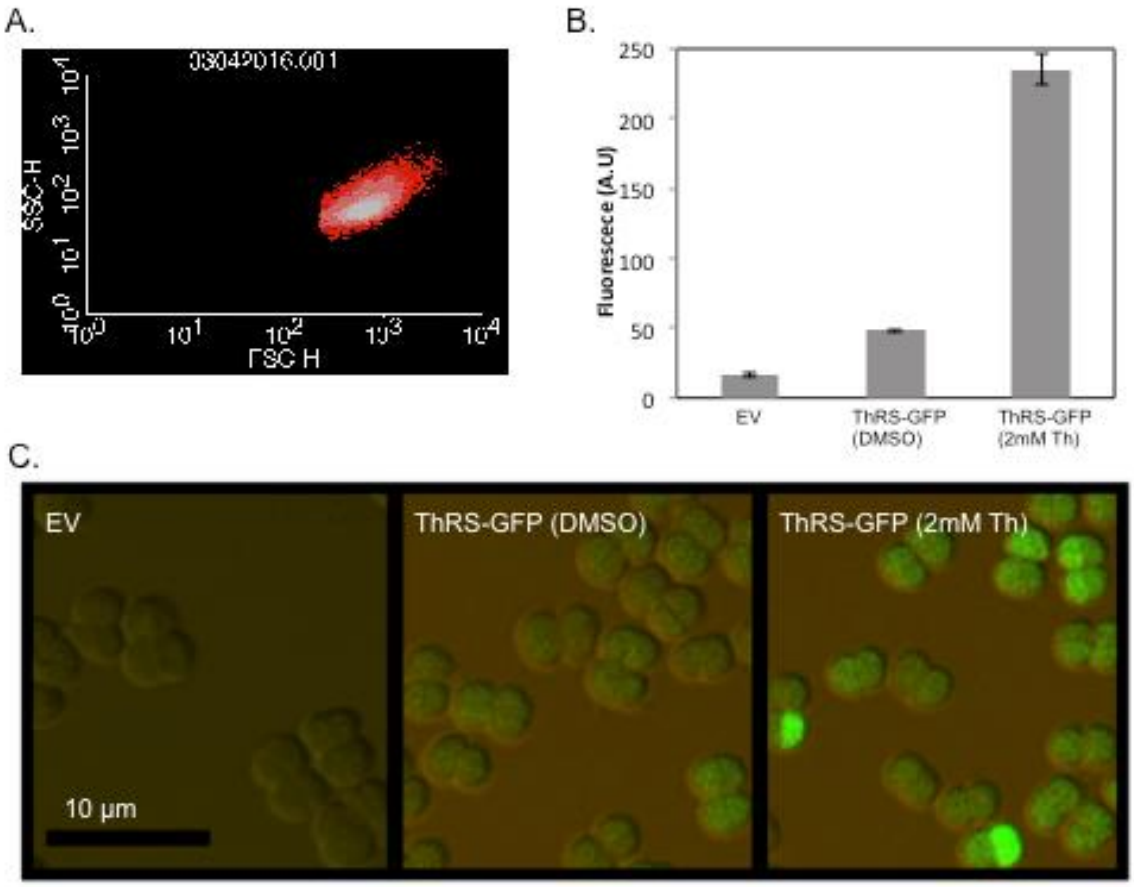

Figure 2.4: *Detecting GFP induced by theophylline riboswitch*. A, FSC v SSC plot to determine localization of *D.radiodurans* population for fluorescence reads. B, Average fluorescence values for testing theophylline induction, EV= empty pRadGro negative control, ThRS-GFP (DMSO) = pRadGro-ThRS-GFP construct tested only with DMSO, ThRS-GFP (2 mM Th) = pRadGro-ThRS-GFP construct induced with 2 mM theophylline. C, Fluorescent microscopy revealing increased GFP signal for theophylline-induced pRadGro-ThRS-GFP strain.

The 41 experimentally confirmed 5'UTR candidate sequences were used to construct a library of potential regulatory regions in the context of the GFP reporter above. Each construct contained the 5'UTR region as well as 60 additional nucleotides corresponding to the nearby coding region (as shown in 1Table 2.6); in this way, these constructs encoded for 20 additional amino acids of the native gene as we reasoned that nucleotides in the coding region may affect the structure of 5' UTR which has been found in other regulatory systems in other bacteria (65). Following construction of this screening assay, the 5'UTR-GFP constructs were screened for increased fluorescence from possible 5'UTR regulatory responses to irradiation.

In vivo fluorescence-based screening reveals potential 5'UTR candidates that regulate gene expression in response to ionizing radiation. The 41 5'UTR candidates uncovered in the bioinformatics approach were tested for regulation with our *in vivo* fluorescence screen under acute (10 kGy) ionizing radiation. It was reasoned that acute doses closer to the survival threshold of *D. radiodurans* (28) would elicit a stronger 5'UTR activation, facilitating their detection. For these assays, *D. radiodurans* R1 samples expressing the pRadGro-5'UTR-GFP reporter were irradiated with a linear accelerator (LINAC) β-ray source. Survival curves were created under these conditions to verify that the expected stress was being imposed to the samples during irradiation, demonstrating that IR stress imposed by LINAC was comparable to previous reports in the literature (28). Following irradiation, 5'UTR activation was observed by fluorescence cytometry and measured by shifts in GFP fluorescence, as any activation of the 5'UTR would result in a downstream increase in expression of the GFP reporter. The GroES promoter of the pRadGro plasmid does not appear to respond to radiation, so it was

concluded that any changes in GFP expression are due to the regulation of the 5'UTR rather than any plasmid-based artifact (56).

For ease of labeling, each 5' UTR candidate tested was referred to by its respective Genscript ID (simple numeric code consisting of odd numbers ranging from 3-83,featured in Table 2.6). Upon visual inspection of the flow cytometry plots, it appeared a few 5' UTR candidates produced a shift upon irradiation (UTRs: 37 and 61, corresponding to DR1913 and DR1279). However, once all of the replicates were consolidated and compared to shifts for the appropriate controls (empty vector negative control, pRadGro, and constitutive non-inducible, pRadGro-GFP), it appeared only one candidate was significantly upregulated (p value < .005), 5'UTR 37 corresponding to the DNA gyraseA gene (DR1913). However, Both of these genes have been observed to be upregulated in response to IR in previous proteomic studies (56). To further examine the impact of radiation or threshold of activation, the screen was also conducted at 1 kGy. The screen was conducted at the same facility with the same LINAC source. Initial analysis of the flow cytometry plots gave the impression that 5' UTR candidates 19 and 37 were both upregulated post-IR. However, more rigorous statistical analysis determined only 5'UTR 37 (gyrA) was significantly upregualted, as before, but this time to a lesser extent. These findings are represented below, Figure 2.5.

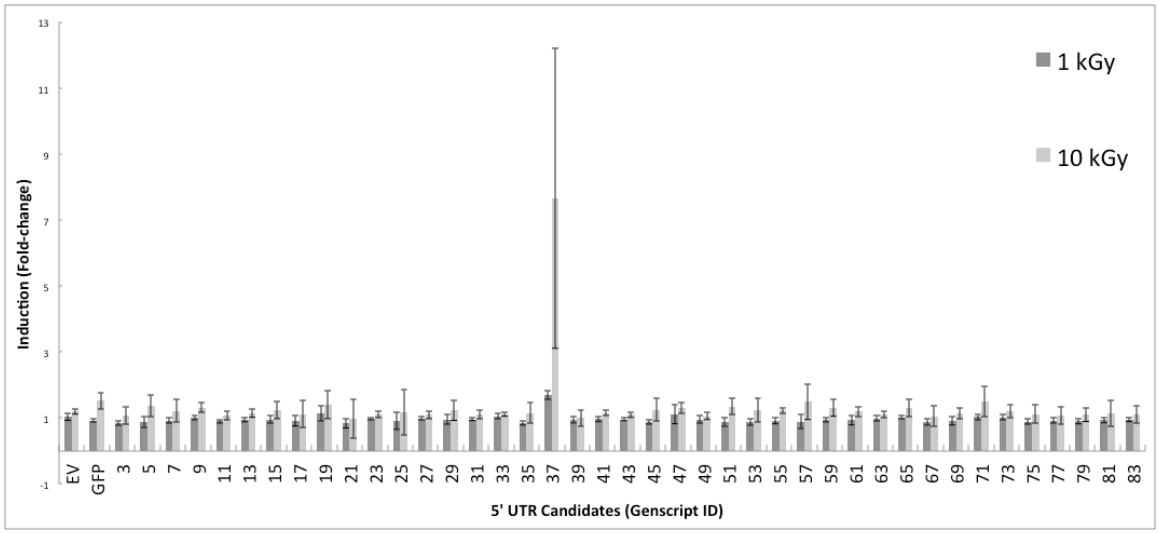

Figure 2.5: *Fluorescent induction of 5' UTRs post-IR treatment*. Fold-change is plotted for all 5'UTR candidates and control, EV = empty vector, GFP = pRadGro-GFP non-inducible control, for 1 and 10 kGy.

To further examine expression dynamics in these test strains, Northern blots were conducted to assay GFP transcript levels upon irradiation for both 1,1, and 10 kGy for all three 5' UTR candidates that initially showed some degree of upregulation. Notably, the GyraseA UTR (5'UTR 37), which was the only 5' UTR candidate to exhibit statistically significant upregulation, showed an approximate 1.5 fold increase in transcription between 1 and 10 kGy. This modest increase in transcription may indicate that the greater increase in fold-change observed at the translational level (via flow cytometry) between 1 and 10 kGy over 4.6 fold may result predominantly from post-transcriptional 5' UTR regulation, Figure 2.6.

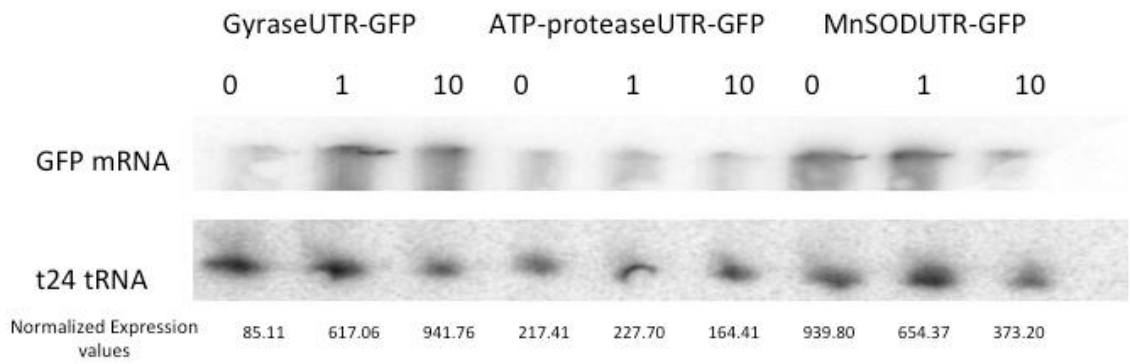

Figure 2.6: *Northern blot of GFP for potentially upregulated 5'UTRs*. Northern blots for 5' UTRs 37, 19, and 61 corresponding to the genes Gyrase subunit A, ATP-dependent protease, and Mn-family Superoxide, respectively, for dosages of 0,1, and 10 kGy. Expression levels were normalized to blot for tRNA 24 (t24) expression as loading control.

In order to corroborate the translational level increases observed by flow cytometry, Western blots for GFP were also conducted for all three of the 5' UTRs that appeared to be upregulated (Figure 2.7). Indeed, these blots confirm the increase of GFP translation, which is most notable in for GyraseA 5' UTR. The observed increase is also greater than the transcriptional increase presented in the Northern blots (Figure 2.6), which, again, could potentially indicate 5' UTR regulation is contributing to GFP expression.

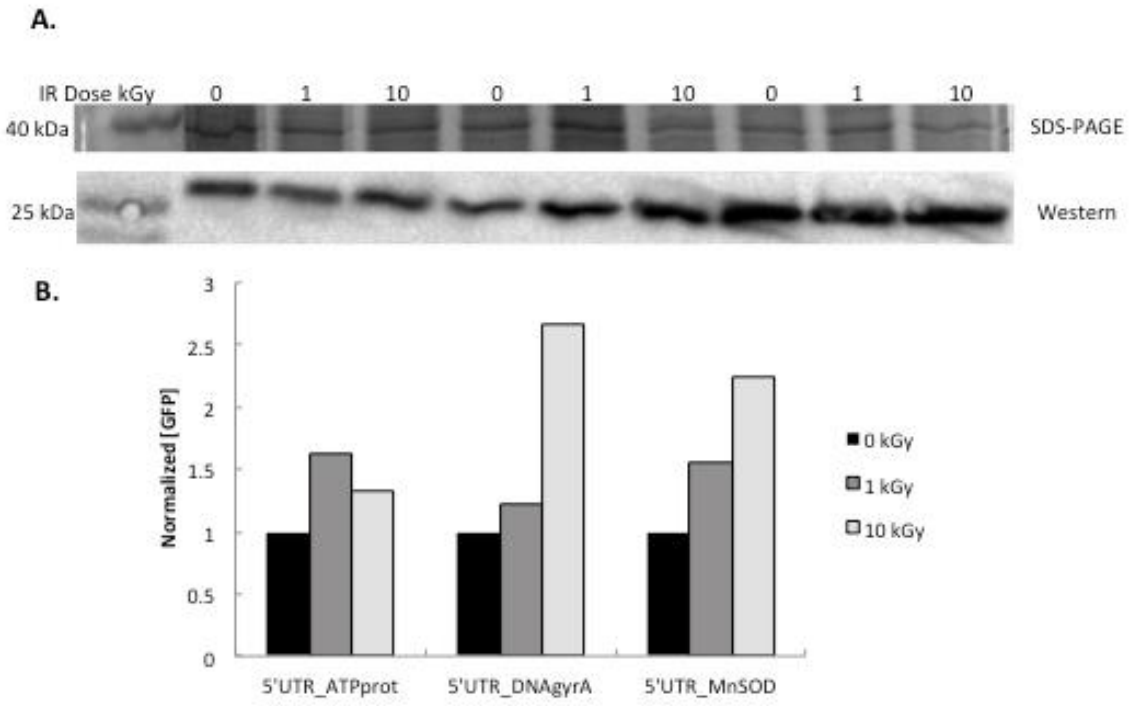

Figure 2.7: *Western blot of GFP for potentially upregulated 5'UTRs*. A, Western blot for GFP under 0,1, and 10 kGy normalized to a prominent 40 kDa band (SDS-PAGE) as loading control. B, Plot of normalized GFP expression levels from Western blot.

DNA GyraseA 5' UTR exhibits IR dose-dependent behavior. The IR treatments at 1 and 10 kGy significantly activated GFP expression for the GyraseA 5' UTR. Moreover, the amount of GFP activation exhibited at the higher dosage also corresponded to a much higher increase in activation. Thus, it was hypothesized that this 5' UTR activates gene expression in a dose-dependent manner. To test for this behavior, the fluorescence screen was conducted as before with the dosages 1, 5, 10, and 14 kGy. Additionally, lower-dose responses were collected using a Cs-137 gamma source to

achieve lower dosages of 100, 250, and 500 Gy. These screens broadly indicate that the magnitude of GFP activation is dependent on dosage, though it appears to be less pronounced for dosages above 5 kGy, Figure 2.8.

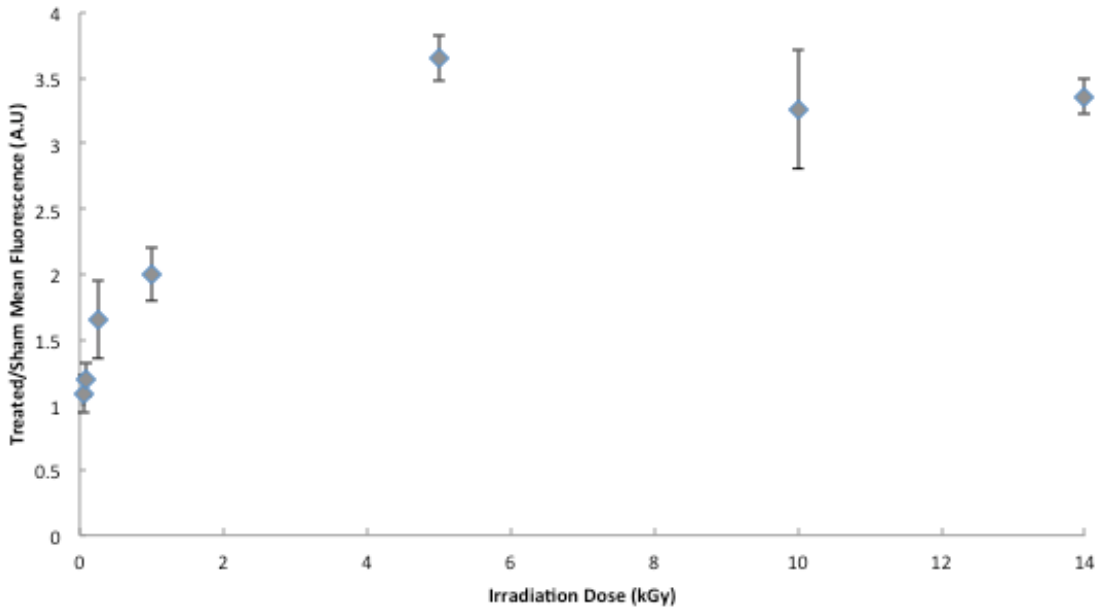

Figure 2.8: *Dosage-dependent behavior of GyraseA 5'UTR*. Induction of GFP for GyraseA 5' UTR generally increases as IR dosage increase. This is especially prominent at lower dosage regime (less than 5 kGy).

In vivo fluorescent-based screen under H_2O_2-induced oxidative stress demonstrates that 5'UTR gene activation is specific to recovery from IR while uncovering a H_2O_2 responsive 5'UTR. While ionizing radiation causes damaging double stranded breaks, additional damage occurs due to the production of reactive oxygen species (ROS) from the irradiation-based cleavage of water (28). Given the possibility that the increase in reporter fluorescence observed during post-IR recovery for the three 5' UTR candidates (5' UTRs, 19, 37, 61) is part of a generic oxidative stress response, rather than a specific IR response, the screen was conducted again as before with H_2O_2 stress instead of IR. Moreover, the screen was conducted with the full library of 41 5' UTRs to identify any candidates that may present a H_2O_2 specific response. This assay was conducted with 15mM H_2O_2 using a Catalase knockout strain (KatA- strain) that is more sensitive to H_2O_2 to amplify the effects of oxidative stress (44). The majority of the 5' UTR candidates did not present any increase in GFP activation (most actually exhibited a decrease in GFP signal). The increased GFP expression observed for the three 5' UTRs upon IR stress was not observed under the H_2O_2 stress. However, 5' UTR 71, corresponding to the annotated organic hydroperoxide resistance protein (OHRP), did increase GFP activation in this screen. While the increase is not overtly pronounced it is significant (p-value < .005) and noteworthy, as there exists a generally negative trend for all the other candidates. Moreover, the increase observed for this 5' UTR was consistent every time the experiment was reproduced, Figure 2.9.

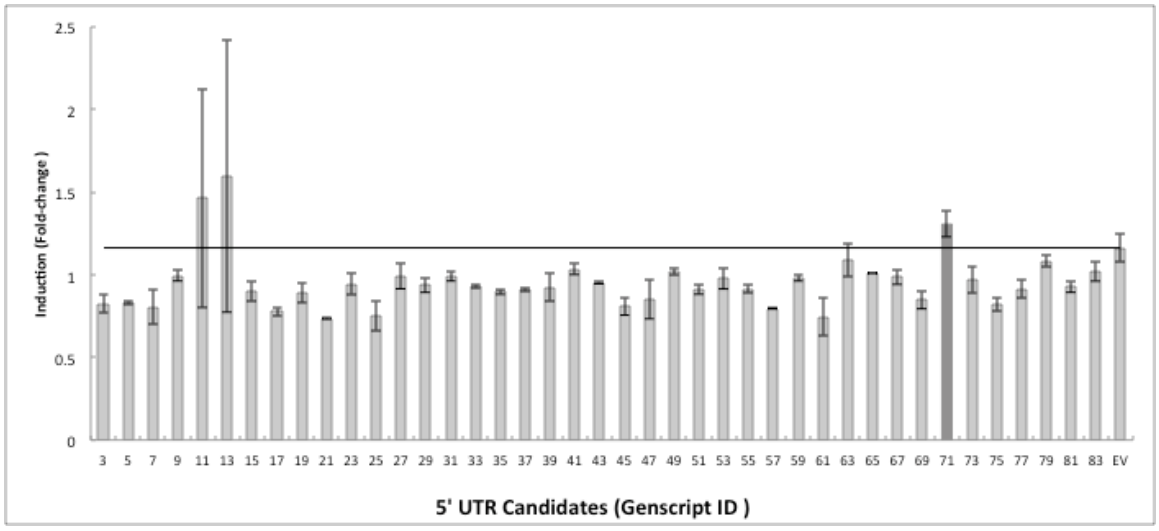

Figure 2.9: *Fluorescent induction of 5' UTRs post H_2O_2.* Induction plotted for all 5' UTRs, EV = empty pRadGro vector control. Bar for 5' UTR 71 is highlighted (dark grey), as it is the only significant increase observed.

Following this original screen, the OHRP 5' UTR was singled out for testing under different concentrations of H_2O_2. While there was no discernable trend towards dose-dependence behavior, the GFP was consistently and significantly activated, Figure 2.10.

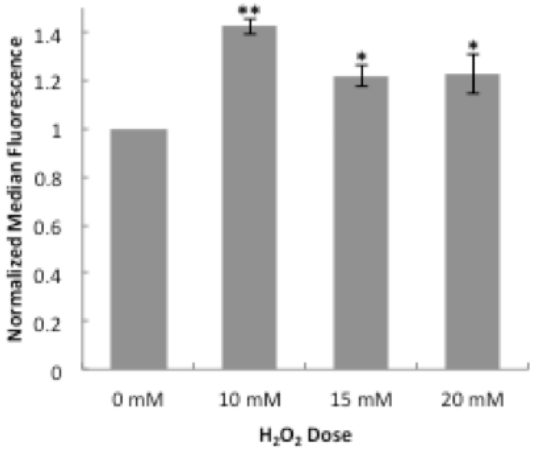

Figure 2.10: *Fluorescent induction of OHRP 5' UTR under H_2O_2 treatment.* Fluorescence plotted for OHRP 5' UTR at 0,10,15,and 20 mM H_2O_2. Double asterisks denote significantly higher fluorescence than the 0 mM control with p value < .005, single asterisk is p value < .05.

Initial Genetic studies portend critical contribution of the 5'UTR_DNAGyr for IR stress recovery. To evaluate the physiological relevance of the regulatory 5'UTR regions within their native genomic context, 5'UTR disruption strain for each candidate of interest were constructed. Specifically, 5' UTRs corresponding to the genes OHRP (DR1857ΔUTR), GyraseA (DR1913ΔUTR), Mn-family superoxide dismutase (DR1279ΔUTR), and ATP-dependent protease (DR349ΔUTR), where each respective 5'UTR region was disrupted in the same manner (as detailed in the methodology) in the context of the WT *D. radiodurans R1* strain. Initially, survivability of these strains post-1, 5, 10 and 15 kGy IR was tested and found limited decreases in survival for the DR1279ΔUTR and DR349ΔUTR strains in comparison to the WT R1 survival fraction. The most significant decrease in IR survivability was observed in the DR1913ΔUTR strain within a regime of 1-10 kGy, Figure 2.11. At this point the data seemed to indicate

that the 5'UTR of DNA gyrase A plays an important role in IR stress recovery in *D. radiodurans*. Unfortunately, subsequent mass spectrometry analysis conducted to probe protein levels of the 5' UTR candidates' downstream genes later revealed that the UTR disruption method employed severely reduced native expression levels for the Gyrase A 5' UTR mutant, Figure 2.12. Thus, no conclusive result could be drawn from this initial genetic analysis to assert the importance of the Gyrase 5' UTR for regulation of the GyrA protein.

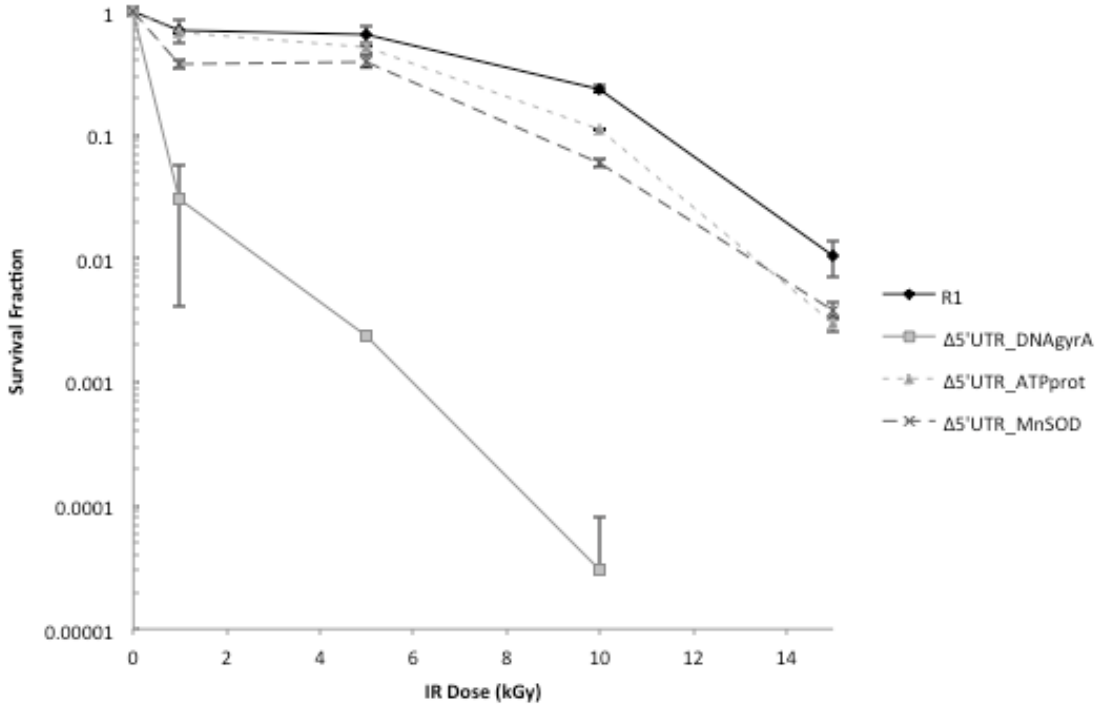

Figure 2.11: *Survival curves of 5'UTR mutants under IR*. Wild-type and 5'UTR mutants of *D.radiodurans* treated with 0,1,5,10 and 15 kGy of IR and plated for survival.

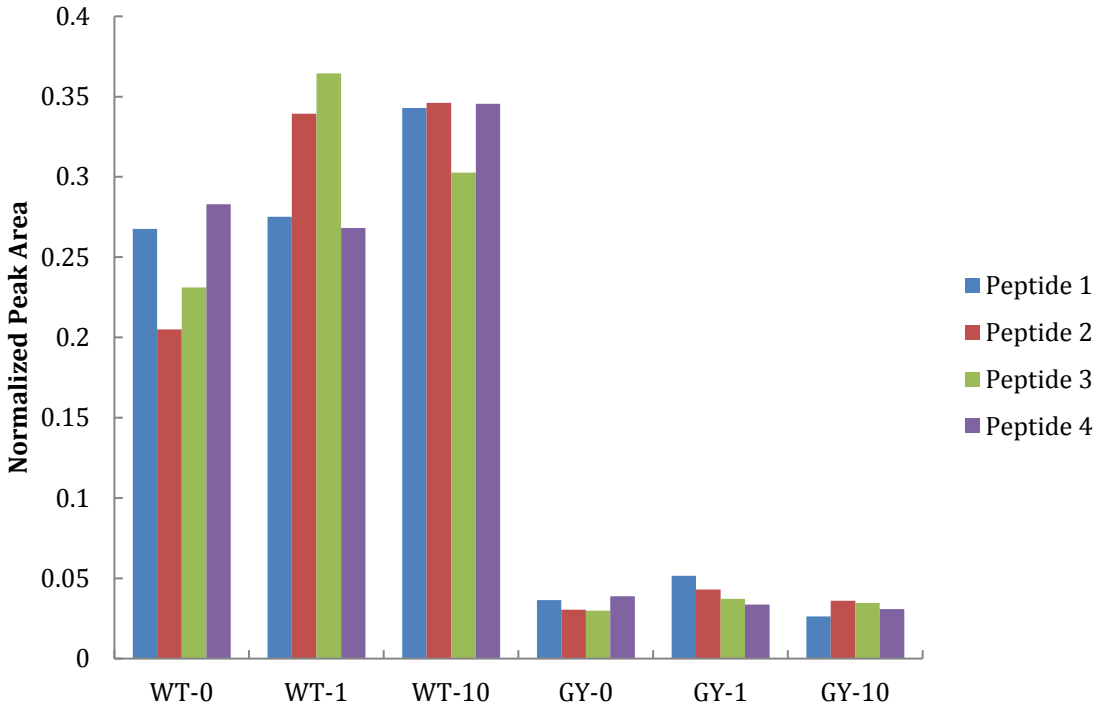

Figure 2.12: *Mass spec for GyraseA expression.* Wild-type (WT) and Gyrase 5' UTR mutant (GY) expression levels for gyrase A protein based on detection four different peptides.

Similarly, survival was tested for the ORHP 5' UTR mutant with a range of H_2O_2 concentrations. This strain was tested along with the WT *D.radiodurans R1* and the Catalase deficient variant, KatA- strain. However, no discernable trend or decrease was observed for the UTR disruption, Figure 2.13. This result might be expected as the increase or activation, though significant, was not very pronounced and might not be detectable with this survival assay.

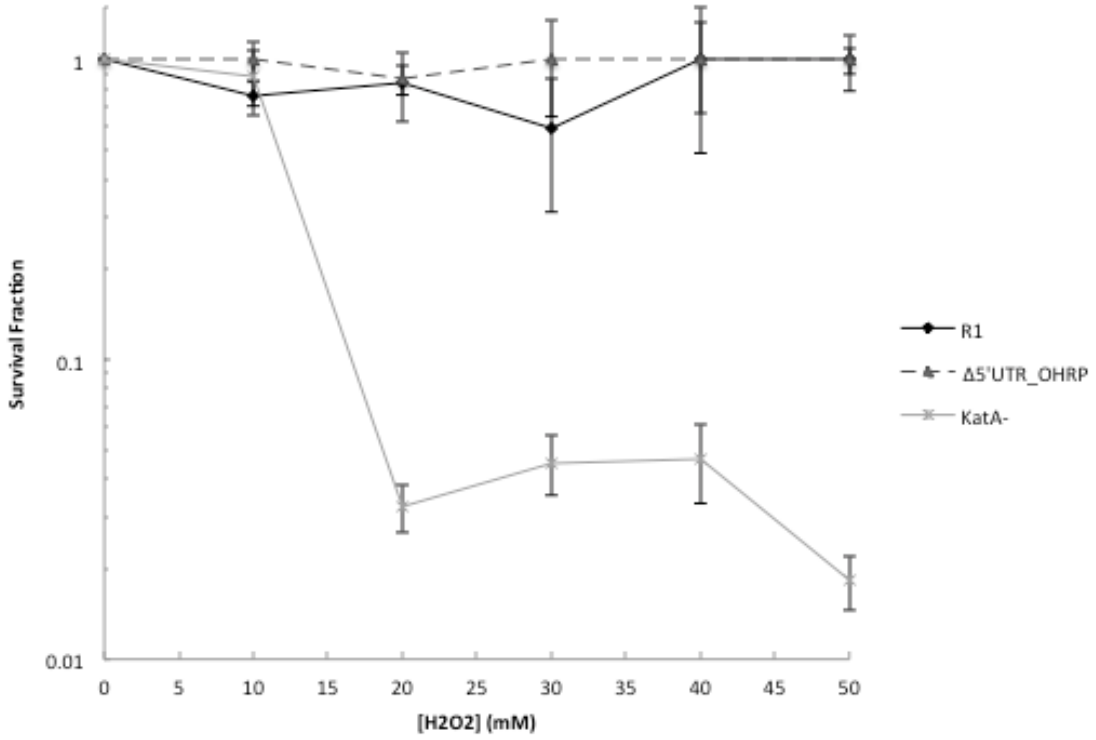

Figure 2.13: *Survival curve of OHRP 5'UTR mutant under H_2O_2 treatment*. Survival of OHRP 5' UTR muatant as compared to wild-type (R1) and the Catalase mutant (KatA-) for H_2O_2 concentrations of 0,10,20,30,40, and 50 mM.

Mechanistic characterization of the gyrase A 5' UTR reveals DdrO dependent regulation at the post-transcriptional level. Identification of the gyrase A 5'UTR through 5'RACE analysis revealed a previously described promoter-based motif present in the 5'UTR of DNA gyrase A, that has been implicated the radiation and desiccation response (RDR) in *D. radiodurans and D. geothermalis* (17). The RDR regulon contains ~29 bioinformatically predicted genes in *D. radiodurans* that are important for survival to radiation, many of which are DNA repair enzymes. The XRE (Xenobiotic Response Element) family transcriptional regulator family repressor protein, DdrO (DR2574), binds specifically to the RDR motif (RDRM) to repress these genes (43). A protease, IrrE (also referred to as PprI) (DR0167), site-specially cleaves the 23 C-terminal residues of DdrO following irradiation (42, 43, 59, 66). This theoretically prohibits DdrO dimerization and encourages release of the RDRM by lowering DdrO affinity to DNA (43). The mechanism of IrrE activation is still unknown, however it is hypothesized to be some type of conformational change as IrrE is constitutively expressed (17). Previous studies have described the RDRM present in the promoter region of the regulon genes, however our 5'RACE data demonstrates the RDRM presence in the 5'UTR of two RDR genes in our candidate list (DNA gyrase A and UvrA).

To elucidate if the radiation dependent regulation observed in the *in vivo* fluorescence screen was based on DdrO repression of the RDRM contained in the GyrA 5'UTR GFP reporter, we repeated the same IR assay in a strain lacking IrrE (ΔIrrE)(34), Figure 2.14. If the reporter response is controlled by IrrE/DdrO in the RDR, then no increase in fluorescence should be observed post-IR as without IrrE being expressed, DdrO will not be derepressed post-IR and activation of fluorescence will not occur. These results demonstrated no increase in fluorescence post-IR, concluding that Gyrase 5' UTR response is largely based upon the RDRM contained in the 5'UTR of GyrA.

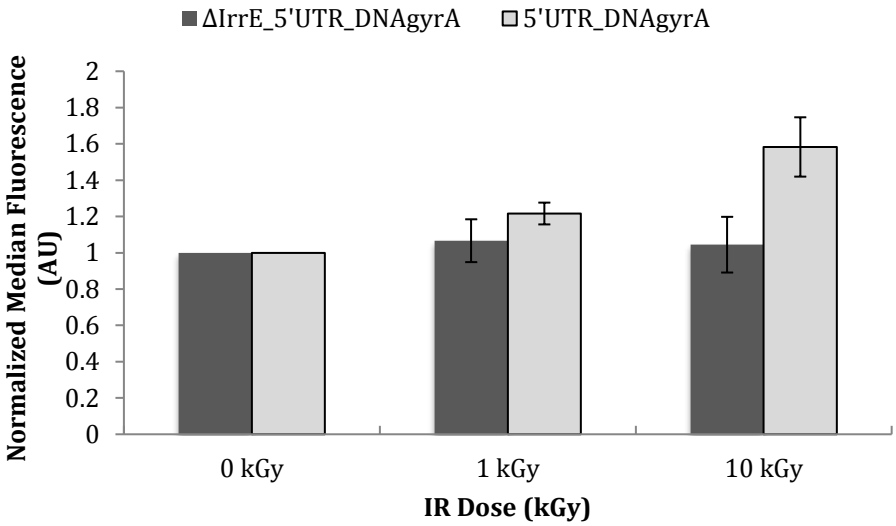

Figure 2.14: *Fluorescent induction of Gyrase 5' UTR in ΔIrrE strain post-IR treatment*. Fluorescent reporter for the gyrase 5' UTR in the IrrE KO (dark grey) strain compared to WT(Light grey).

Furthermore, because of the presence of the RDRM in the 5'UTR, instead of the promoter like previously thought, additional experimentation was necessitated to determine if DdrO could also bind RNA. DdrO contains a putative N-terminal Cro/C1-type helix-turn-helix domain present in the XRE (Xenobiotic Response Element) family transcriptional regulators (43). While this family of transcriptional regulators primarily binds DNA, it could be possible to bind RNA as a subfamily motif, a winged helix-turn-helix, can bind RNA (67). To determine if DdrO demonstrated the ability to regulate RNA as well as DNA in the 5'UTR of Gyrase A, a gel mobility shift assay (GMSA) of the DNA (positive control) and RNA samples containing the 5'UTR of GyrA with DdrO was conducted. DdrO was found to be able to bind RNA as well as DNA, albeit with

different affinities, Figure 2.15. This finding suggests that there are multiple layers of regulation of the RDR, controlled by the 5'UTR of the genes in the regulon**.** Notably, the other RDRM containing gene (UvrA, DR1771) in this study did not show any fluorescence response to irradiation, suggesting that there could be different stresses that trigger this response.

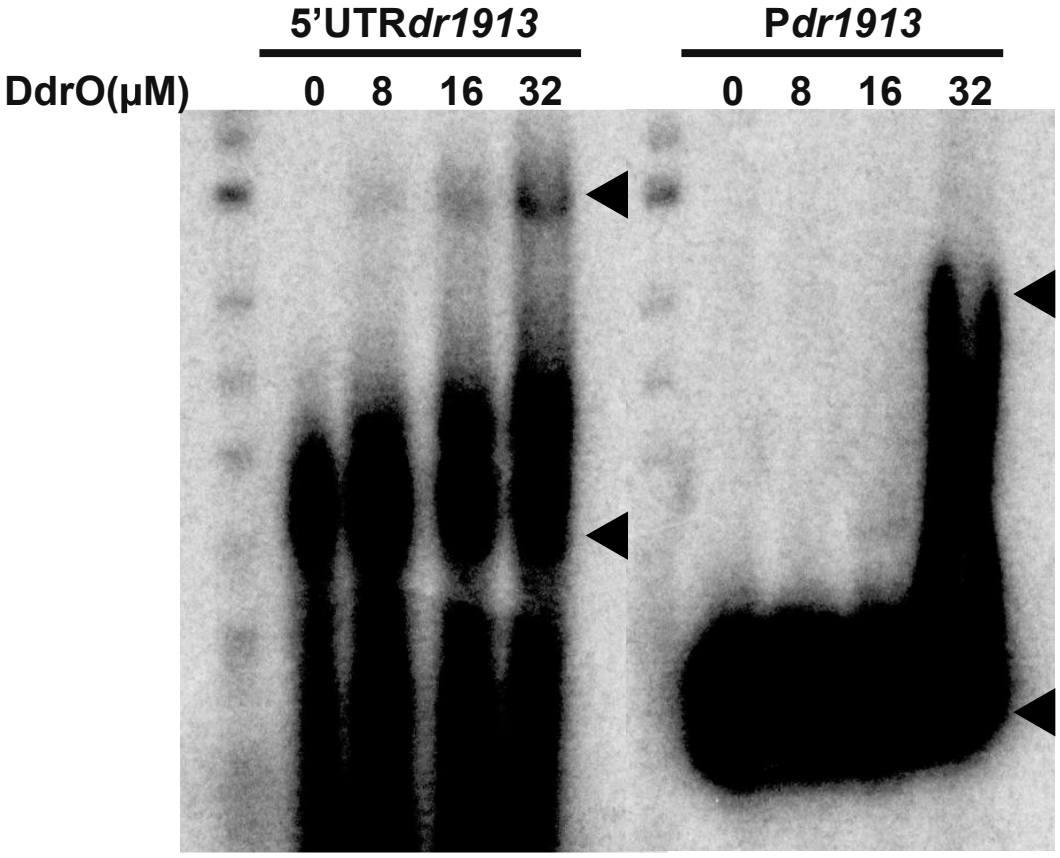

Figure 2.15: *GMSA for DdrO binding RDRM*. Phosphor image of radiolabeled RNA (5'UTRdr1913) and DNA (Pdr1913) corresponding to 5' UTR of gyrase harboring RDRM. Arrows denote location of bands corresponding to a smaller band (non-bound) and larger band (bound to DdrO). Concentrations of DdrO used are indicated above image.

DISCUSSION

This work capitalizes on sequencing data much like other contemporary works. However, where the majority of sequencing-based projects primarily make use of genetic annotations or transcription levels, this work seeks to elucidate non-coding regions that may be harder to detect within the transcriptome. The general dearth of knowledge and annotations for UTR regulatory elements provides a challenge that is further compounded by studying more novel organisms. Previous studies in *D.radiodurans* have analyzed sequencing data to mine the transcriptome for regulatory RNA elements such as radiation responsive sRNAs and leaderless mRNAs (38, 68). The results presented in this work comprise initial efforts to systematically identify putative 5' UTRs that upregulate gene expression under stress. The process could probably be further optimized or even automated to generate a list of candidate UTRs without manual analysis of the sequencing data. The manual analysis, visually inspecting transcriptome data on IGV (Integrated Genome Viewer) for UTRs, was a given constraint to the throughput of this work. It is possible there are yet regulatory 5' UTRs that were not covered in this investigation. The downstream bioinformatics and experimental pipeline aimed to refine the UTR search by consolidating structural, conservation, and expression level data with experimental verification to guide our functional screening of gene activation. This type of approach could be adapted for other organism and may be especially useful for less characterized organisms, such as extremophiles that could reveal novel RNA function.

The limited results obtained in this work include some interesting findings. While the only statistically significant post-IR increase in expression was observed for the gyrase 5' UTR, two other candidates exhibited more subtle increases that might have suffered from detection limits imposed by the nature of experimentation. These potentially activated candidate 5' UTRs correspond to the genes for ATP-dependent

protease (DR349) and Mn-family superoxide dismutase (DR1279). Both these genes were initially selected for further testing because they produced a very slight but consistent shift observed in the fluorescent cytometry plots. However, both were revealed insignificant when more rigorous statistical analyses were performed. However, as previous work has revealed, both of these genes have been observed to be upregulated upon irradiation (56). The ATP-dependent protease, also known as Lon2, makes sense as an important node of regulation. Previous work has observed that the *Deinococcus* genome harbors a large number of potential proteases (69), many of which are present upon gamma irradiation during the recovery period (70, 71). In addition, experiments testing ATP-dependent protease knockouts revealed their critical nature for recovering from IR and puromyocin stress (72). Indeed, proteases seem to be generally necessary for clearing damaged proteins and aiding recovery. Thus, it stands to reason organisms might develop UTR-based regulation for rapid and robust response and recovery. Similarly, the Mn-family superoxide dismutase has been determined to be important for post-IR recovery (44). Notably, the Mn-SOD is the only active SOD in *D.radiodurans* and represents an important avenue for ameliorating oxidative stress (73).

However, the standout result from this work has been the observed activation of gene expression under control the gyrase A 5' UTR. As previously mentioned, this gene is part of the RDR system (17). Currently, this regulatory system is understood to operate uniquely at the promoter level where the regulatory protein, DdrO, binds and inhibits radiation-response genes that are activated upon cleavage by the protease IrrE (34). Activation of this protease is thought to occur upon irradiation, though the mechanism is not fully elucidated. As the premise of this work proposes post-transcriptional regulation at the 5' UTR as a powerful mechanism for rapid response to stress, the 5' UTR of gyrase A was further characterized to uncover any UTR-based regulation. The initial UTR-

knockout strategy that was implemented was not successful at elucidating post-transcriptional functionality. However, further assays utilizing a strain devoid of an important RDR component, IrrE, suggests the 5' UTR is regulated by the RDR system. However, this assay did not necessarily test whether the RDR regulation was occurring at the DNA or RNA level. Previous work determined binding of DdrO to various targets of the RDR system (including gyrase A) at the DNA level (42). However, the IR assays revealed activation levels of GFP increasing with increasing dosage. If this regulatory response is limited to DNA, in our experimental case, regulation would be limited to acting on a low-copy plasmid (45, 63) that might essentially produce a more muted binary response. The stark increase between the activation observed at .1 kGy compared to 10 kGy suggests there could be additional layers of regulation. Moreover, the palindromic nature of the RDRM binding site could also intimate a conserved RNA stem loop secondary structure that facilitates protein binding (refs). Thus, the GMSA conducted in this work presents critical corroboratory evidence of RNA-based regulation. This is especially salient as many of the predicted RDRM binding sites (17) are present further downstream than any annotated promoter regions and likely in the UTR. The major contribution of this work has been the expanded model of the RDR regulatory system that present UTR-level regulation as the most efficient route for gyrase A induction (and perhaps RDR genes not tested in this work) within this system, Figure 2.16.

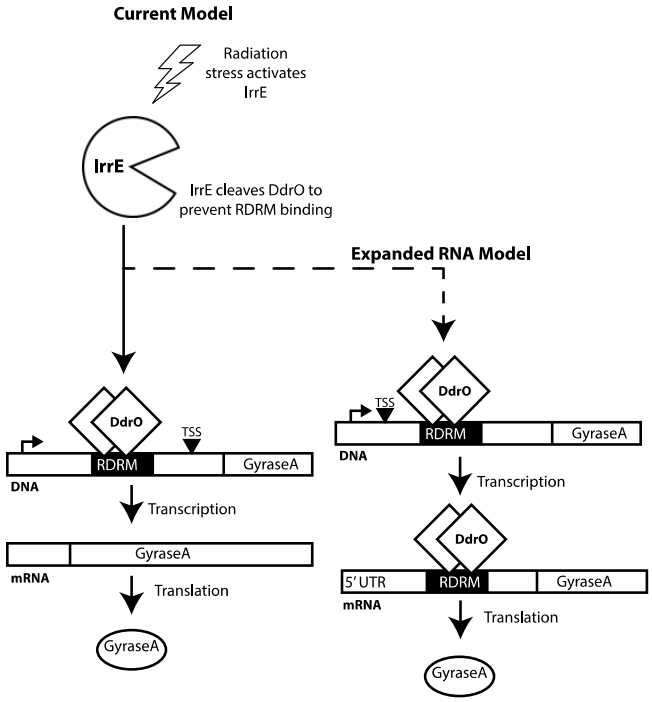

Figure 2.16: *Expanded model for RDRM-based induction of GyrA*. This model proposes an additional route of regulation upon irradiation and activation of the IrrE protease. Current model posits transcriptional activation mediated by cleavage of DdrO at RDRM upstream of transcriptional start site (TSS). Expanded model places RDRM in UTR downstream of TSS, where abundant mRNA can be regulated for exponentially robust protein expression.

The overarching aim of this work was to uncover UTR-level regulatory elements that advance the utility of RNA-based regulation. While the experiments have had limited success thus far, the expanded model of the RDR system serves as a starting point. Additionally, the experiments in this work also revealed one 5' UTR candidate, the largely uncharacterized OHRP gene, as potentially sensitive to H_2O_2 oxidative stress.

This candidate still requires more experimental verification and was beyond the immediate scope of this work. However, these results help to sustain the potential for discovery of novel RNA elements that may be exploited for bioengineering purposes.

Chapter 3: Rational Engineering of Regulatory ncRNA

INTRODUCTION

The discovery of noncoding small RNAs (sRNAs) as regulators has led to an entire new class of metabolic regulation that has already found a significant place in synthetic biology and metabolic engineering (14, 39, 74, 75). Noncoding sRNAs often serve various functions within cells, contributing to the regulation of critical phenotypes such as the ones involved during recovery from environmental stresses, virulence, and even plasmid conjugation (40, 75). Mechanistically, noncoding sRNAs can elicit a regulatory response by base-pairing and sequestering the Shine-Dalgarno (SD) sequence of the ribosome-binding site or the start codon of a target mRNA. These sRNAs can occur as trans-encoded species that are coded elsewhere on the genome relative to the target gene, in many bacterial species relying on interactions with the, Hfq protein chaperone to facilitate base-pairing (7, 39). There are also cis-encoded sRNAs that are coded on the anti-sense strand of target genes and base-pair with the untranslated regions (UTR) or terminators to form RNA duplexes that interfere with translation, termination, or mRNA stability (76). In addition to their ability to interact with mRNAs, sRNAs can also play significant roles by sequestering regulatory proteins (40, 77).

Additionally, sRNAs can also play larger regulatory roles by sequestering regulatory proteins. One such sRNA that interacts with a regulatory protein is *csrB*. The sRNA, *csrB*, is part of the carbon storage regulator (Csr) system, which relies on the regulatory protein CsrA to bind a variety of mRNAs to elicit a change on expression. The specific function *csrB* serves within the csr system is that of a "sponge" that binds multiple (approximately 18-22) CsrA protein homodimers to counter-act their regulatory effects (78, 79). Together, the sRNA and protein elements regulate the expression of

hundreds of different genes in *E.coli* that encompass a large range of function (80). Thus, the Csr system is an important regulatory system that contributes to many of the complex phenotypes that may be worth "tuning" such as virulence and metabolism. Previous efforts to exploit Csr regulation have relied on *csrB* overexpressions to modulate the effects of CsrA regulation and produce levels of metabolites and amino acids desirable for production of biosynthetic commodities (15). However, overexpression strategies often present biological systems with an undue metabolic load and can be taxing to the innate physiology of the cell. This is especially true when exploiting global regulatory systems such as the Csr system. And while it may be desirable and convenient to exploit an sRNA-protein interface responsible for the regulation of many different genes, more efficient strategies are necessary to fully capitalize on the potential of the Csr system.

A novel approach for tuning the regulatory capacity of sRNAs, such as *csrB*, may be to alter the sites on the sRNA that interact with regulatory elements to modulate binding and regulatory capacity. Within the Csr system, *csrB* is primarily interacting with the CsrA protein. In fact, CsrA-binding motifs have been previously defined and essentially consist of stem-loops that contain GGA sequences in the loop and allow for predicting potential binding sites (18). Moreover, three-dimensional structures have been determined for homologs of *csrB* in *Pseudomonas fluorescens* that map out the exact binding sites of proteins (81). However, there exist no structural determinations of *E.coli* that can reveal the exact binding sites of CsrA or other proteins on *csrB*. In this work, a previously developed RNA structure-sensing reporter was utilized for assessing the accessibility of *csrB* to identify sites of lower accessibility that may correspond to sites that bind regulatory elements. The *in vivo* RNA Structural Sensing System (iRS^3) utilizes an mRNA that contains a region complementary to an sRNA of interest. If the probing region of the mRNA binds to the sRNA, then, the RBS in the mRNA becomes accessible

to allow translation of GFP (82). Notably, upon probing the entire sequence of *csrB*, an accessibility profile was generated that revealed differential accessibility for the predicted CsrA-binding sites. The regions potentially binding CsrA were then assessed individually to determine the contribution to CsrA regulation using a fluorescent reporter (83). These efforts informed the modular design of engineered *csrB* species with varying levels of Csr regulation. This work represents a novel approach to engineering sRNAs, and in doing so, presents a global regulatory system that is amenable to "fine tuning" to mediate the expression of many genes.

METHODOLOGY

In vivo RNA Structural Sensing System

For the determination of potential CsrA binding sites along *csrB* the previously developed iRS^3 method was employed to generate a base-pairing accessibility profile (Steve Jorge 2013). This assay utilizes an mRNA that consists of a GFP coding sequence and an upstream "cis-blocking" (CB) stem loop region that sequesters a RBS. The functional component of this system is a probe sequence fused to the 5' end of the mRNA that is free to bind target RNA molecules. If there is successful base pairing with the probe sequence, then the CB stem loop structure is disrupted and permits binding of the ribosome to the RBS to initiate translation of the GFP. Expression of this system is driven by a pZER21αγ12aG plasmid that includes a pLtetO promoter to drive expression of the iRS^3 construct and a pBAD promoter to express the target RNA. A schematic of the iRS^3 is illustrated below in Figure 3.1.

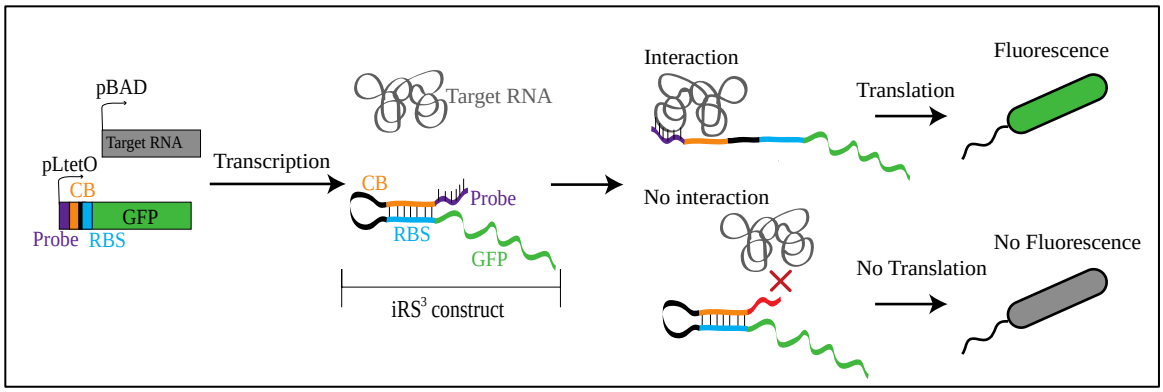

Figure 3.1: *Fundamentals of iRS*^{3} (adapted from Sowa et al. 2014). The plasmid drives transcription of target RNA and the iRS^3 construct. If the iRS^3 probe binds a the target RNA, a conformational change is induced that opens the CB loop and enables translation initiation for the production of the GFP signal.

Strains and Plasmid Preparation

The pZER21αγ12aG-iRS^3 plasmid was prepared for assaying accessibility of *csrB* by restriction cloning of the *csrB* sequence, as PCR-amplified from wild-type *E.coli* K-12 MG1655, downstream of the arabinose-inducible pBAD promoter using SalI and XbaI cut sites (explained in Table 3.1). The original pZER21αγ12aG-iRS^3 plasmid conferred kanamycin resistance, which was potentially incompatible with some of the *E.coli* testing strains, so the KanR cassette was swapped for an AmpR cassette from a pBAD-18 plasmid using Gibson Assembly (84) to confer ampicillin resistance. All strains and plasmids are included in Table 3.1.

Strain	Description	Notes	Source
DH5a	*E.coli* K-12MG1655 derivative,F-φ80*lacZΔM15* Δ(*lacZYA-argF)U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-*	Commercial competent cell strain used for cloning	Invitrogen™
CML381	*E.coli* K-12MG1655 derivative,Δ*csrB*Δ*csrC*	Used for iRS^3 accessibility assay	Tony Romeo Lab
HL4142	*E.coli* K-12MG1655 derivative,*lacI$_q$* Δ*csrA*Δ*csrB*Δ*csrC*Δ*csrD*Δ*glgCAP*Δ*pgaABCD*	Initially used for testing CsrA regulation assay	Adamson and Lim, 2013
HL3712	*E.coli* K-12MG1655 derivative,lacI$_q$ Δ*csrB*Δ*csrC*Δ*glgCAP*Δ*pgaABCD*	Used for data collection for CsrA regulation assay	Adamson and Lim, 2013
Plasmids			
pZER21αγ12aG-iRS^3	ColE1 origin, pLtetO-iRS(3), pBAD, KanR	Parent plasmid for iRS^3	Lab Stock
pZER-iRS^3-csrB	ColE1 origin, pLtetO-iRS(3), pBAD-csrB, KanR	Parent plasmid for iRS^3 with csrB inserted; Restriction cloning SalI and XbaI, insert forward primer: GAGCCATATGACCGTCGACAGGGAG TCAGAC insert reverse primer: TCCGCTTCTAGAAATAAAAAAAGGG AGCACTGT	Lab Stock
pZER-iRS^3-csrB-AmpR	ColE1 origin, pLtetO-iRS(3), pBAD, AmpR	pZER-iRS^3-csrB with KanR substituted for AmpR; Gibson Assembly primers: *Vector* [Forward:ACATATTTGAATGTATTTTT GATGCCTGGCAGTTCCCTAC Reverse:TCACTGATTAAGCATTGGTAA GCGGGACTCTGGGGTTCGA], *Insert* [Forward:GTAGGGAACTGCCAGGCAT CAAAAATACATTCAAATATGTATCCG CTCATGAG Reverse:TCGAACCCCAGAGTCCCGCTT ACCAATGCTTAATCAGTGAGGC]	Lab Stock
pBAD-18	pBR322 origin, pBAD promoter, Amp R	Used as template for amplifying AmpR cassette used to clone pZER-iRS^3-csrB-AmpR	Lab Stock
pHL600	p15a origin, pLtetO-csrB, pLlacO-csrA, KanR	Expresses CsrA and *csrB* for CsrA regulation assays	Adamson and Lim, 2013
pHL1756	ColE1 origin, pConM12-glgC UTR-gfp, pCon-tetR, AmpR	Expresses *glgC 5'* UTR fused to gfp for reporter in CsrA regulation assay	Adamson and Lim, 2013

Table 3.1: *Strains and plasmids*

Probe Design

Probe sequences were designed according to previously published specifications (82) to cover the majority of the *csrB* molecule, excluding the Rho-independent terminator region, and cloned directly upstream of the CB region of the iRS^3 construct under control of the pLtetO promoter using Gibson Assembly (primers listed in Table 3.2). A collection of 27 pZER-iRS^3-csrB-AmpR plasmids was generated, each coding for expression of a different probe sequence, and transformed into a Δ*csrB*Δ*csrC* K-12 MG1655 strain (CML 381, Table 3.1).

Probe	Length	Target Start	Target End	Sequence	Forward Primer	Reverse Primer
1	10	4	13	ACUCCCUGUC	gaattcACTCCCTGTCtaccattcacctcttggatttg	tggtaGACAGGGAGTgaattcggtcagtgcgt
2	13	15	27	CACUUCGUUGUCU	gaattcCACTTCGTTGTCTtaccattcacctcttggattt	gaattcCACTTCGTTGTCTtaccattcacctcttggattt
3	14	32	45	UGUCAUCAUCCUGA	cgaattcTGTCATCATCCTGAtaccattcacctcttggattt	TCAGGATGATGACAgaattcggtcagtgcgtc
4	12	46	57	GUCCUGCAGAAG	cgaattcGTCCTGCAGAAGtaccattcacctcttggattt	CTTCTGCAGGACgaattcggtcagtgcgtc
5	14	57	70	ACCAUCCUGGUGUG	gaattcACCATCCTGGTGTGtaccattcacctcttggattt	ggtaCACACCAGGATGGTgaattcggtcagtgcgtc
6	11	73	83	CUUUCCCUGAA	gaattcCTTTCCCTGAAtaccattcacctcttggatttg	tggtaTTCAGGGAAAGgaattcggtcagtgcgt
7	12	85	96	UUCAUCCAGAAG	gaattcTTCATCCAGAAGtaccattcacctcttggatttg	tggtaCTTCTGGATGAAgaattcggtcagtgcgt
8	13	99	111	CGUCAUCCUCUUC	cgaattc CGTCATCCTCTTC taccattcacctcttggattt	GAAGAGGATGACGgaattcggtcagtgcgtc
9	11	109	119	GCGUCCUGCGU	cgaattcGCGTCCTGCGTtaccattcacctcttggattt	ACGCAGGACGCgaattcggtcagtgcgtc
10	11	122	132	GGUGUCCUUUA	cgaattcGGTGTCCTTTAtaccattcacctcttggattt	TAAAGGACACCgaattcggtcagtgcgtc
11	12	135	146	UUCUCCAUCCUG	gaattcTTCTCCATCCTGtaccattcacctcttggattt	ggtaCAGGATGGAGAAgaattcggtcagtgcgtc
12	11	147	157	ACCGGUUCUCA	cgaattcACCGGTTCTCAtaccattcacctcttggattt	TGAGAACCGGTgaattcggtcagtgcgtc
13	11	154	164	CAUCCUGACCG	cgaattcCATCCTGACCGtaccattcacctcttggattt	CGGTCAGGATGgaattcggtcagtgcgtc
14	11	166	176	GACCCACCGAA	aattcGACCCACCGAAtaccattcacctcttggatttg	ggtaTTCGGTGGGTCgaattcggtcagtgcgtc
15	11	176	186	UGGCCUUCCUG	cgaattcTGGCCTTCCTGtaccattcacctcttggattt	CAGGAAGGCCAgaattcggtcagtgcgtc
16	12	184	195	AAGUGUCCCUGG	gaattcAAGTGTCCCTGGtaccattcacctcttggattt	ggtaCCAGGGACACTTgaattcggtcagtgcgtc
17	13	193	205	CUUCAUCCUGAAG	gaattcCTTCATCCTGAAGtaccattcacctcttggattt	ggtaCTTCAGGATGAAGgaattcggtcagtgcgtc
18	11	212	222	ACCACCCCGAU	cgaattcACCACCCCGATtaccattcacctcttggattt	ATCGGGGTGGTgaattcggtcagtgcgtc
19	11	229	239	AUUGCUUCCUG	cgaattcATTGCTTCCTGtaccattcacctcttggattt	CAGGAAGCAATgaattcggtcagtgcgtc
20	12	244	255	UCGUUCAUCCUG	cgaattcTCGTTCATCCTGtaccattcacctcttggattt	CAGGATGAACGAgaattcggtcagtgcgtc
21	12	255	266	CUUGCGGCCAAU	gaattcCTTGCGGCCAATtaccattcacctcttggattt	ggtaATTGGCCGCAAGgaattcggtcagtgcgtc
22	10	267	275	UUCCUCUGGC	gaattcTTCCTCTGGCtaccattcacctcttggattt	ggtaGCCAGAGGAAgaattcggtcagtgcgtc
23	11	270	280	AACUUUUCCUC	cgaattcAACTTTTCCTCtaccattcacctcttggattt	GAGGAAAAGTTgaattcggtcagtgcgtc
24	11	281	291	UCAUCCUUGAC	gaattcTCATCCTTGACtaccattcacctcttggattt	ggtaGTCAAGGATGAgaattcggtcagtgcgtc
25	12	294	305	UUGUUGCUCCCUG	gaattcTTGTTGCTCCTGtaccattcacctcttggattt	ggtaCAGGAGCAACAAgaattcggtcagtgcgtc
26	12	310	321	AGCAUUCCAGCU	cgaattcAGCATTCCAGCTtaccattcacctcttggattt	AGCTGGAATGCTgaattcggtcagtgcgtc
27	12	324	335	CCGGUUCGUUUC	cgaattcCCGGTTCGTTTCtaccattcacctcttggattt	GAAACGAACCGGgaattcggtcagtgcgtc

Table 3.2: *RNA probes for iRS*[3]

Fluorescent cytometry for generating accessibility profile

The iRS[3] experiments were conducted using fluorescence cytometry and triplicate cultures of each probe (with *csrB* induced or non-induced) to measure the fluorescent output. To this end, cultures of CML381 containing pZER-iRS[3]-csrB-AmpR plasmids were grown overnight in Luria-Bertani (LB) media at 37° C. The next day the saturated overnight cultures were used to seed fresh ten ml cultures at 1:100 dilution in LB media containing 50 μg/ml of ampicillin and allowed to grow for two hours at 37° C. At this point, cultures were either induced for *csrB* expression with 100 ng/μl of arabinose or non-induced for each probe. All strains were supplemented with 100 ng/μl of anhydrotetracycline (aTc) to ensure expression of the iRS[3] cassette under the pLtetO, which should nonetheless be constitutive (as no Tet repressor was coded for in plasmid or genome). The cultures were prepared for fluorescence cytometry at four hours post-induction by pelleting 100 μl of each sample and re-suspending in one ml of 1X phosphate buffered saline. These samples were then run in a Becton Dickinson FACScalibur cell sorter with a 488 nm argon laser and a 530 nm FL1 logarithmic amplifier and measured for fluorescence. At least 150,000 cells were counted for each sample. Values of accessibility for each probe were determined as the log of the average ratio of median fluorescence (in arbitrary units) of induced to non-induced samples for each replicate and error was calculated as the standard deviation of the induced to non-induced ratios among replicates.

Mutations of csrB for determination of CsrA-binding ability

Mutations of *csrB* were informed by the iRS^3 accessibility profile, structural predictions, and literature. Previous work had identified general characteristics of CsrA binding sites; GGA motifs have been observed in transcripts that are bound by CsrA (85). Additionally, *in vitro* SELEX experiments determined CsrA had high affinity for stem loop structures with GGA motifs in the loop region (18). Therefore, regions of *csrB* predicted to have stem loop structures (Mfold structural prediction, (86)) with GGA in the loop region that also exhibited relatively low accessibility (presumably due to CsrA binding preventing base pairing of iRS^3 probe) were selected as candidates for mutagenesis. As the aim of mutagenesis was to test CsrA binding for specific regions of *csrB*, variants of the RNA molecule were designed to abrogate CsrA binding without altering the predicted structure by either single nucleotide substitutions or modular substitution of the entire predicted stem loop region with that of the high accessibility (presumably due to increased base pairing of iRS^3 probe due to absence of CsrA binding) stem loop. Single nucleotide substitutions were executed as transversions of the GGA motif to GCA, as this was previously observed to greatly reduce affinity for CsrA (18). Structural predictions of the mutants and wild-type *csrB* molecules were generated using Mfold to avoid designing mutants that introduced structural changes, as this may be a confounding variable in testing CsrA binding. These initial mutants were cloned using Gibson Assembly with the primers in Table 3.3.

csrB mutant	Description	Forward Primer	Reverse Primer
5 SNP	GGA motif in loop corresponding to probe 5, mutated to GCA	CACACCAGCATGGTGTTTC AGGGAAAGGCTTCT	AACACCATGCTGGTGTGTCCTG CAGAAGTGT
5:18ST	Stem loop corresponding to probe 5, substituted with stem loop corresponding to probe 18	ACACATCGGGGTGGTGTTT CAGGGAAAGGCTTCT	AACACCACCCCGATGTGTCCTG CAGAAGTGTCATCA
7 SNP	GGA motif in loop corresponding to probe 7, mutated to GCA	AGGCTTCTGCATGAAGCGA AGAGGATGACGCA	TCTTCGCTTCATGCAGAAGCCTT TCCCTGAAACACC
7:18ST	Stem loop corresponding to probe75, substituted with stem loop corresponding to probe 18	CATCGGGGTGGTGCGAAGA GGATGACGCAGG	TCGCACCACCCCGATGCCTTTCC CTGAAACACCATCC
22 SNP1	First GGA motif in region corresponding to probe 22, mutated to GCA	ATTGGCGGCAACGCCAGAG GAAAAGTTGTCAAGGAT	TCTGGCGTTGCCGCCAATCGTTC ATCCTGAACTATTGCTT
22 SNP2	Second GGA motif in loop corresponding to probe 22, mutated to GCA	GCCAGAGCAAAAGTTGTCA AGGATGAGCAGGGA	CTTGACAACTTTTGCTCTGGCCT TGCGGCCAAT
25:18ST	Stem loop corresponding to probe 25, substituted with stem loop corresponding to probe 18	TGAGCGGGGTGCAACAAAA GTAGCTGGAATGCTG	TTTGTTGCACCCCGCTCATCCTT GACAACTTTTCCTCTG

Table 3.3: *Initial single region csrB mutants*

Fluorescent reporter for measuring CsrA binding

The CsrA binding ability of the mutagenized *csrB* molecules was tested using a previously developed GFP plasmid reporter system (83). This reporter system consists of two plasmids; plasmid pHL600 drives expression of CsrA and *csrB* using pLlacO (IPTG inducible) and pLtetO (aTc inducible), respectively, plasmid pHL1756 constitutively expresses a GFP mRNA with the 5' UTR of a well-characterized CsrA-binding transcript, *glgC* (83, 87). As illustrated below in Figure 3.2, the reporter system functions by expressing a GFP mRNA that is translationally repressed by CsrA binding and activated by *csrB* sequestration of CsrA. Thus, expression of *csrB* mutants produces differential levels of GFP signal corresponding to CsrA sequestration and binding ability relative to a wild-type *csrB*.

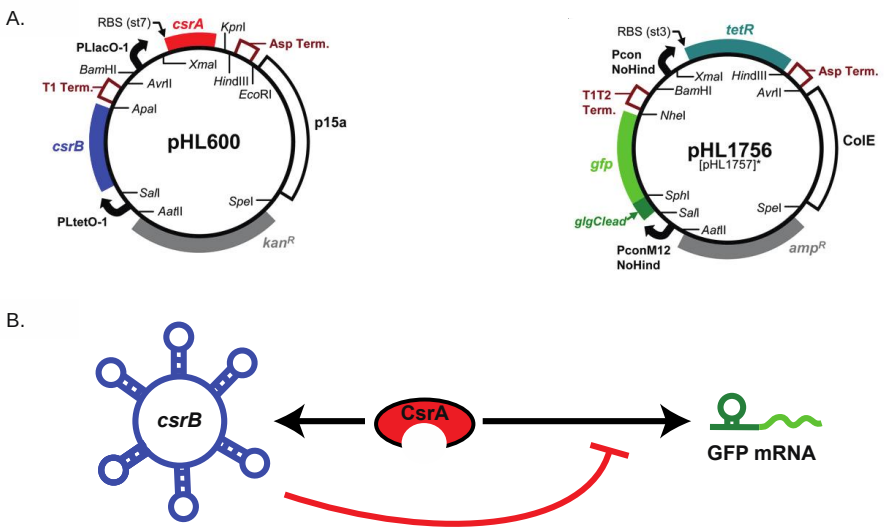

Figure 3.2: *CsrA-binding fluorescent reporter system* (as adapted from Adamson and Lim 2013). A, plasmid system, B, interactions of CsrA and *csrB* regulate GFP expression.

To test for differential CsrA binding across *csrB* mutants the reporter plasmids were transformed into the HL3721 strain (Δ*csrB*Δ*csrC* K-12 MG1655 derivative). This strain yielded much clearer and consistent results than a previously tested strain, HL4142 (Δ*csrA*Δ*csrB*Δ*csrC*Δ*csrD* K-12 MG1655 derivative, data not shown). Fluorescence output was measured as before with the Becton Dickinson FACScalibur, collecting data from at least 100,000 cells. The experimental procedure consisted of growing overnight cultures of sample strains each expressing different *csrB* mutants along with the wild-type, all in biological triplicate in LB media at 37° C. The next morning, 100 µl of the saturated overnight cultures were used to seed ten ml cultures in LB media with kanamycin (pHL600, 50 µg/ml) and ampicillin (pHL1756, 50 µg/ml), each biological replicate seeding two cultures (for induced and non-induced samples). The samples were allowed to grow for approximately 2 hours until they reached an optical density at 600 nm (OD_{600}) of .4 - .6. Upon reaching the desired OD_{600}, the samples were either induced with aTc (final concentration, 1 µM) to drive *csrB* expression, or non-induced to serve as controls for the effect of each *csrB* mutant on CsrA regulation. The CsrA cassette in pHL600 was not induced, as the strain used, HL3721, produces sufficient CsrA at native levels. The samples were then prepared for fluorescence cytometry as before at two hours post-induction. This time point was experimentally determined to yield the clearest results across mutants. The extent of CsrA regulation of the GFP reporter (as mediated by *csrB* sequestration) was calculated as the average of the ratio of the mean fluorescence of each induced *csrB* sample to its corresponding non-induced *csrB* biological replicate for each *csrB* mutant. Error was calculated as the standard deviation between the *csrB* induced/non-induced ratios for each replicate of a mutant. Mutants of *csrB* were determined to have significantly different levels of CsrA regulation than the wild type by using an unpaired one-tailed T test ($p < 0.05$ or $p < 0.005$).

Generation of combinatorial csrB mutants

Following the initial evaluation of the single-region *csrB* mutants, combinatorial mutants were rationally designed and cloned to produce *csrB* species that exhibit a wider range of CsrA binding. In addition, regions that were observed to exhibit high accessibility and well-defined stem loop structure were also mutagenized by substitution with a low accessibility stem loop with the intent of producing *csrB* species capable of binding more CsrA. As before, discrete predicted stem-loop regions were substituted with stem loops of differential CsrA-binding ability and structural conservation was verified with Mfold predictions. These mutants were then synthesized and cloned into the pHL600 plasmid by GenScript®, Table 3.4. Data for CsrA regulation for these mutants was collected and analyzed as before for the single-region *csrB* mutants.

csrB mutant	Description	Predicted CsrA Affinity
9:25st	csrB probing site 9 substituted with high affinity loop p25	increased
18:25st	csrB probing site 18 substituted with high affinity loop p25	increased
19:25st	csrB probing site 19 substituted with high affinity loop p25	increased
21:25st	csrB probing site 21 substituted with high affinity loop p25	increased
5,7,22,25:18st	csrB probing sites 5,7,22,25 substituted with low affinity loop p18	reduced
5,7,25:18st	csrB probing sites 5,7,25 substituted with low affinity loop p18	reduced
22,25:18st	csrB probing sites 22,25 substituted with low affinity loop p18	reduced
7,25:18st	csrB probing sites 7,25 substituted with low affinity loop p18	reduced
5,25:18st	csrB probing sites 5,25 substituted with low affinity loop p18	reduced
7,22:18st	csrB probing sites 7,22 substituted with low affinity loop p18	reduced
5,7:18st	csrB probing sites 5,7 substituted with low affinity loop p18	reduced
9,18:25st	csrB probing sites 9,18 substituted with high affinity loop p25	increased
9,19:25st	csrB probing sites 9,19 substituted with high affinity loop p25	increased
18,19:25st	csrB probing sites 18,19 substituted with high affinity loop p25	increased
18,21:25st	csrB probing sites 18,21 substituted with high affinity loop p25	increased
9,18,21:25st	csrB probing sites 9,18,21 substituted with high affinity loop p25	increased
9,19,21:25st	csrB probing sites 9,19,21 substituted with high affinity loop p25	increased
9,18,19,21:25st	csrB probing sites 9,18,19,21 substituted with high affinity loop p25	increased

Table 3.4: *Combinatorial csrB mutants*

RESULTS

Accessibility assay reveals potential CsrA-binding sites for mutational analysis

Previous studies established the iRS[3] as an efficient means of determining base-pairing accessibility and intimating structural complexity of non-coding RNA molecules (82). In this work, the utility of iRS[3] was expanded to elucidate protein-binding sites on RNA. The accessibility profile generated for *csrB* with the iRS[3] is displayed below, Figure 3.3.

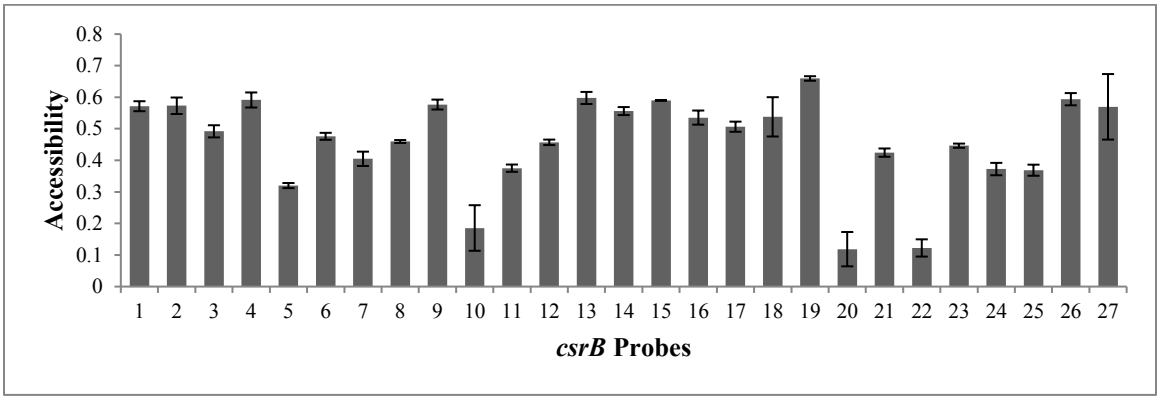

Figure 3.3: *iRS[3] accessibility profile for csrB.* Accessibilty (induced over uninduced fluorescence) plotted for each probe tested.

Many of the sites probed that were predicted to be CsrA binding sites exhibited lower accessibility. Indeed, the five least accessible regions corresponded to the probes 5,10,20,22, and 25, which all contain GGA motifs, as presented in Figure 3.4 (GGA motifs are in bold font). Notably, only one of those regions, corresponding to probe 10, is not predicted to conform to a stem loop with the GGA motif in the loop region. However, the five most accessible regions also include regions that conform to the canonical CsrA-binding sites. Thus further experimentation was prompted to further test the low accessibility sites for CsrA binding.

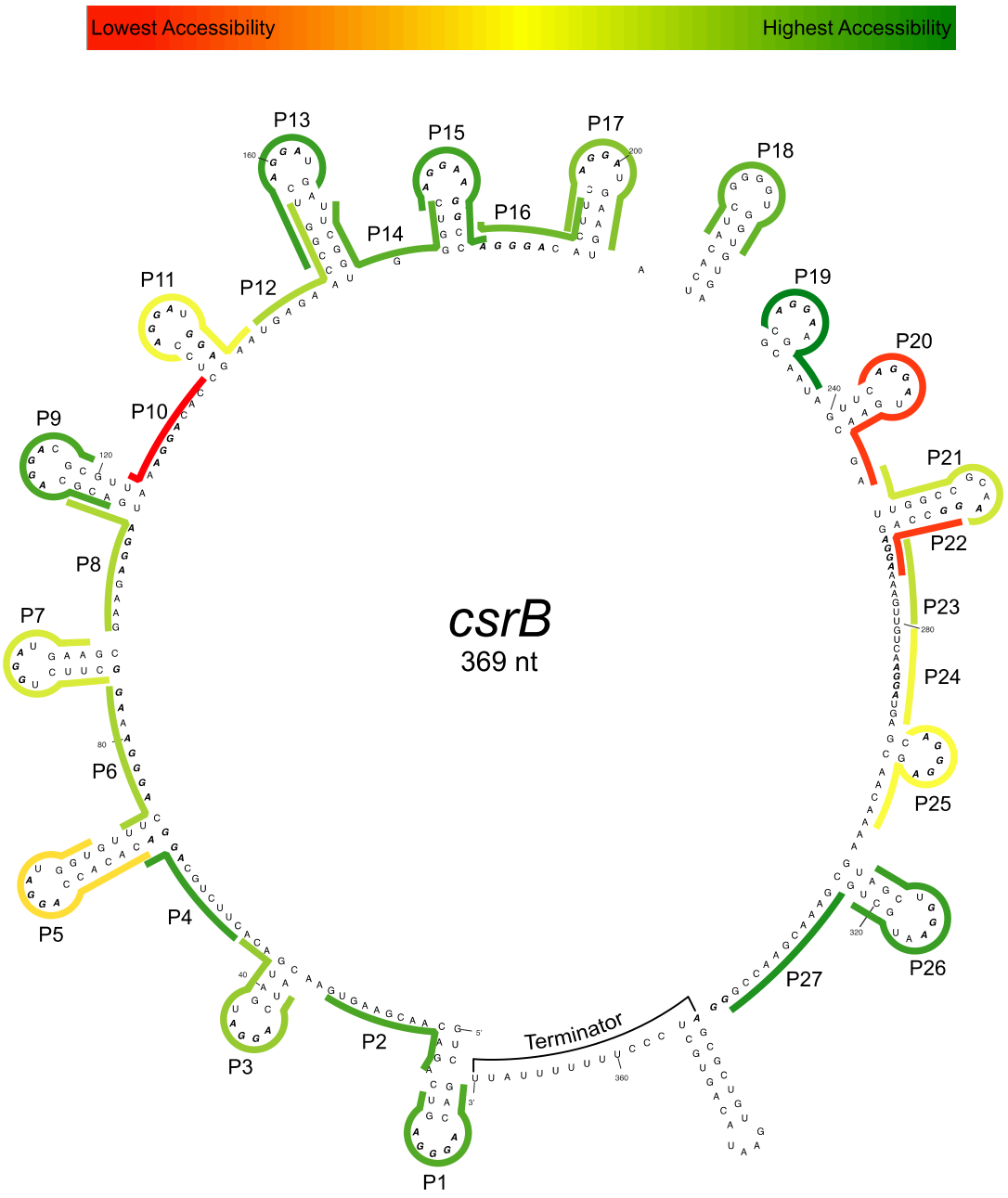

Figure 3.4: *Heat map of iRS[3] probe accessibility for csrB.* Visual representation of individual binding ability, or accessibility, for each probe. Color code above structural prediction of *csrB*.

As a result of the iRS[3] assay, specific sites were identified as putative "strong" CsrA binding regions. Initially, the sites corresponding to probes 5,11,20,22, and 25, which represent the five lowest accessibility regions that are predicted to conform to the canonical CsrA-binding stem loop structure, were selected as candidates for mutational studies. The mutational studies involved the rational redesign of these stem loop sites to non-canonical CsrA binding sites (as explicated in the methodology section) to observe their impact on CsrA binding and regulation. Thus, the candidate regions were analyzed for ease of local mutation and minimal impact to predicted global structure. For this reason, sites corresponding to probes 11 and 20 were not pursued as candidates for mutation and the site corresponding to probe 7 was advanced as a mutant candidate. Moreover, two different mutation strategies were pursued; single nucleotide substitutions involving transversions of GGA motifs to GCA (designated as SNP mutants), and stem loop substitutions involving the modular replacement of the entire stem loop structures with that of the stem loop corresponding to probe 18 (designated as :18ST mutants). The site corresponding to probe 18 was chosen due to the high accessibility and predicted non-CsrA binding stem-loop structure observed that would designate this site as a very unlikely CsrA binding site. Finally, seven single CsrA binding site mutants were designed: 5 SNP, 5:18ST, 7 SNP, 7:18ST, 22 SNP 1, 22 SNP 2, and 25:18ST, as explicated in Table 3.3 of the methodology.

CsrA binding assay refines selection of csrB sites for engineering

This initial set of mutants was tested for their individual impacts on CsrA binding by using a previously developed CsrA regulation assay (83). This assay relies on the inducible expression of CsrA and *csrB* from plasmid pHL600 in trans, and a GFP reporter mRNA (expressed by pHL1756) that is subject to regulation (repression) by

CsrA (see methodology for protocol). These mutants were tested against the wild type *csrB* with the expectation that due to less CsrA affinity, there would be more CsrA available for repression of GFP resulting in a decreased fluorescent output for the mutants, Figure 3.5.

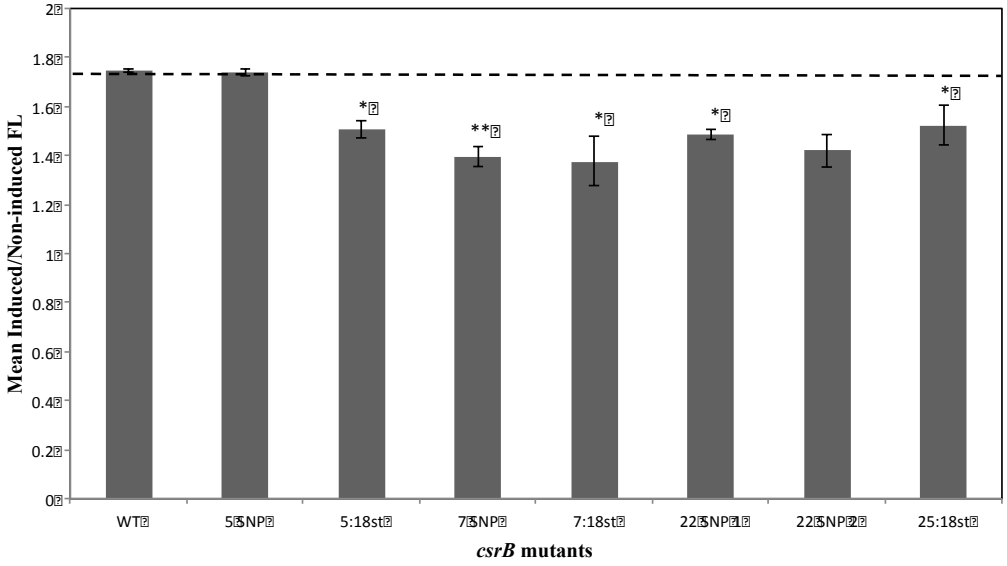

Figure 3.5: *CsrA regulation of csrB mutants*. CsrA regulation (as determined by change in fluorescence) plotted for initial mutants. * = p value < .05; ** = p value < .005

Indeed, the majority of the *csrB* mutants yielded lower fluorescence than the wild type. Only 5 SNP and 22 SNP 2 failed to exhibit significant decreases of fluorescence. The *csrB* mutant, 7:18ST, exhibited the lowest fluorescence, indicating a stronger effect on CsrA regulation. Additionally, among the mutants that had both a SNP and 18ST variant, the stem loop substitution appeared to have the more drastic effect.

Engineered csrB species exhibit gradient of regulatory ability

After identifying and testing *csrB* sites for decreased CsrA binding ability, these sites were employed for designing combinatorial mutations to yield *csrB* species that exhibit a range of CsrA affinity. The mutants were designed by substituting the previously determined CsrA-binding sites, with a low-affinity site (probing site 18) in various permutations intended to produce *csrB* species with lower CsrA affinity than the single-site mutants that had been tested. Moreover, to increase the range CsrA binding affinity, *csrB* species were also designed to increase binding affinity. These variants were designed by substituting low-affinity *csrB* sites with the high-affinity loop corresponding to probing site 25. Furthermore, the combinations were expected to produce variants that exhibited increased CsrA binding. These engineered *csrB* species were tested for CsrA binding and regulation as before and measured with fluorescent cytometry. Indeed, the results of this assay present a wide range of CsrA affinities. With few exceptions the engineered *csrB* variants that were designed and tested generally behaved as expected. The variants designed for low affinity exhibited lower CsrA binding than the wild type *csrB*. However, there was no clear trend demonstrating an additive effect when using variants with multiple substitutions. This is particularly notable for the lower affinity mutants, where *csrB* variants with multiple substitutions do not generally confer a decreased CsrA affinity or regulatory effect. In contrast, the *csrB* variants designed for higher CsrA binding did appear to increase CsrA affinity when multiple low affinity sites were substituted for higher affinity sites. Substituting probing site 21 with a high-affinity loop seems to produce the largest increase in CsrA binding, and combininng this substitution with others only increases that effect, Figure 3.6.

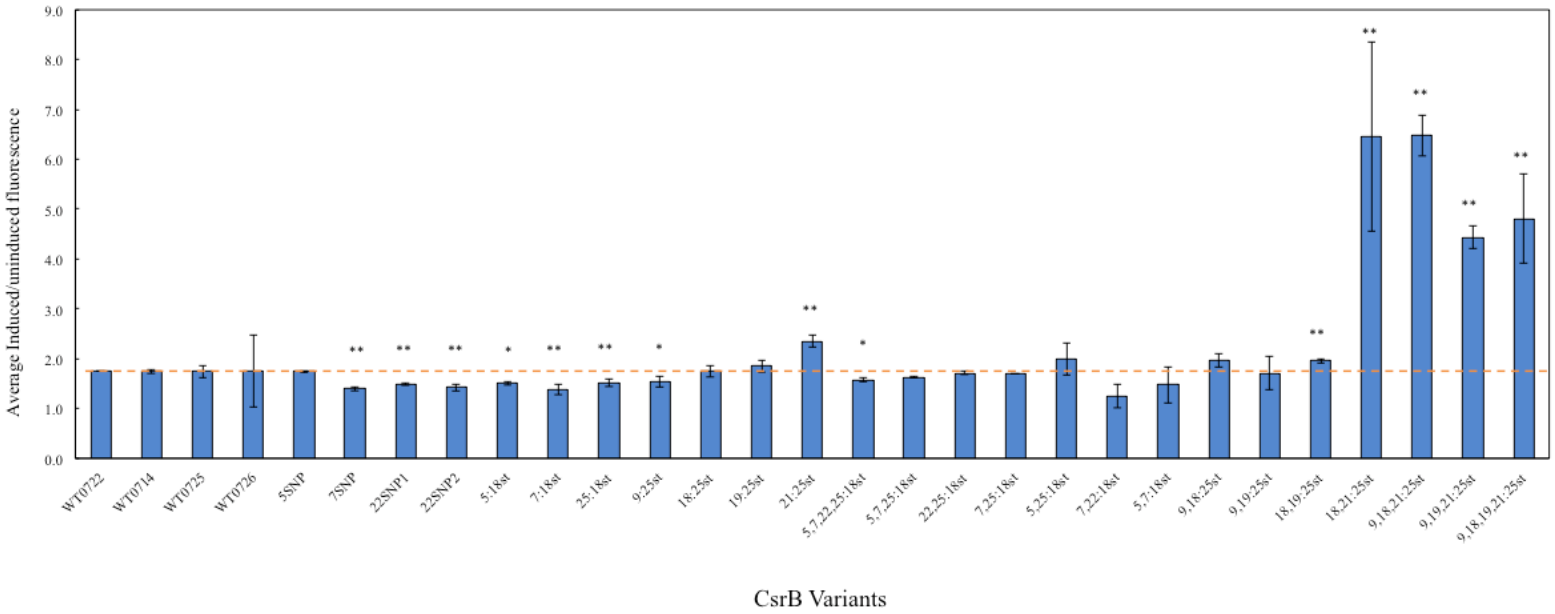

Figure 3.6: *csrB mutants generate gradient of CsrA regulation*. CsrA regulation (as determined by change in fluorescence) plotted for all generated mutants. Each mutant was compared to its respective WT based on date of assay. * = p value < .05; ** = p value < .005

DISCUSSION

The iRS^3 had been previously developed for detecting structure, primarily secondary, on RNA molecules (82). This *in vivo* assay was able to capture dynamic structures by exploiting base-pairing ability of a probe fused to GFP mRNA. In this work, the system was utilized for assessing the binding of proteins to an RNA molecule. Currently, few methodologies have been developed to interrogate protein binding of RNAs *in vivo*. However, there are certain considerations in employing this system for measuring protein binding. Generally, it may be difficult to discern a true protein-binding site from a highly structured region of RNA that abrogates binding of the iRS^3 probe. For this work, the accessibility assays for *csrB* were complemented by previously published findings on the properties of CsrA binding on *csrB* (18, 81, 88–90). Thus, results of the accessibility profile for *csrB* were more likely to reflect true CsrA-binding sites and guided the rational design of CsrA-binding variants. Moreover, there exists no full 3D structure of the *E.coli csrB*, but partial structural characterizations of *csrB* homologs have presented a complex picture of the temporal and spatial dynamics of *csrB* structure (81, 88). However, as this experiment relies on exogenous overexpression of the RNA of interest, *csrB*, and the GFP signal (which served as measure of accessibility) was measured for hundreds of thousands of cells, the accessibility profile that was generated may be representative of the most likely and abundant RNA conformation at that sampling time. As such, the profile generated by this ensemble of data could be paired with the subsequent assays measuring CsrA-binding affinity GFP reporter assay to intimate the most biologically relevant CsrA-*csrB* interactions in the absence of a complete 3D RNA structure.

The sites of *csrB* that were deemed to be important for CsrA binding generally have the structure of canonical CsrA binding sites, that is, a stem loop with GGA motif in the loop. While these findings do not necessarily enhance our understanding of CsrA-*csrB* binding at local interfaces, the global accessibility profile of *csrB* may have revealed the sites that are most critical for this molecule. This approach might be the most direct method utilized for determining functionally important sites in *csrB*. Specifically, the sites corresponding to probes 5,7,22,and 25 exhibited CsrA binding capacity and informed the subsequent engineering efforts. Previous work indicates that CsrA dimers may bind cooperatively by binding pairs of CsrA-binding sites, usually with differential affinity (65). It is not clear from the experiments conducted how the sites that were manipulated affected the cooperative binding of CsrA. Thus, it cannot be confidently determined if the CsrA-binding sites that were selected and mutated were abrogating CsrA binding partially or completely. Similarly, it is difficult to determine how the *csrB* species harboring multiple substitutions are affecting cooperative CsrA binding. For the variants designed to increase CsrA binding, there was a more apparent increase when substituting multiple binding sites. In addition, it is not possible from these experiments to visualize any global changes in structure that occur between the variants upon mutagenizing and the impact on CsrA binding. The intent of this work was to engineer a regulatory ncRNA by substituiting individual bindinung sites in a modular fashion to affect sites locally while prodicing a gloabal effect on CsrA binding with minimal change in structure. For this reason, structural predictions were used in an attempt to minimize any changes n structure that could have unintended implications for binding, thus, undermining the modular nature of the rational engineering methodology.

Ultimately the goal of this work was to produce a library of *csrB* mutants that exhibit a range of CsrA binding affinity to impose a regulatory effect (as reported by

GFP) to modulate phenotypic effects. This methodology does not rely on traditional approaches of tuning transcription or translation rates to control expression levels of genes. Instead of imposing a metabolic and physiological burden on cells by funneling resources to drive expression, the global regulatory ncRNA, *csrB*, was altered to deliver various levels of the genes controlled by the Csr system and tune their expression. Much of this work presents novel elements for imposing control on cells. Aside from the overall approach of targeting a highly conserved bacterial ncRNA for modulating complex phenotypes by exploiting binding sites in a modular fashion, the general concept of manipulating global regulation, such as the Csr system, is still underexplored. Whereas many attempts to affect gene expression engineer promoter strength or RBS affinity, few approaches have exploited native regulatory systems or introduced synthetic regulatory systems to coordinate expression of desirable phenotypes. Some efforts have targeted unique components of large regulatory frameworks, such as transcriptional regulators, small molecules and co-factors (14, 16). Engineering higher-order nodes in native regulatory systems has proven successful in producing high value products, such as aliphatic hydrocarbons, while avoiding the bulky metabolic impositions of traditional metabolic pathway engineering. Oftentimes, the regulatory systems that are engineered have evolved feedback loops and sensors to optimize the balance of synthesis and cell growth (16).

The advantages to targeting multiple genes by manipulating regulatory systems for coordinated remodeling of a cell's physiology have become clear in recent years. Particularly within the context of manipulating the Csr system, where overexpression of *csrB* has produced phenotypes conducive to heterologous expression of valuable products (15). Ultimately, the findings presented in this work represent early efforts to generate modifed *csrB* species in a more efficient rational manner to modulate control over a

global regulatory system. While the architecture of the Csr system may lend itself to the engineering strategy presented in this work, the methodology advanced may be adaptable to other similar regular systems or regulatory RNAs and may serve as the foundation for wholly synthetic regulatory systems.

Chapter 4: Pending Work

In conducting my thesis work, I have developed and performed experiments for two different projects, as presented in chapters two and three. However, both of these works require further experimentation to materialize into peer-reviewed publications.

Genome-wide Screen for Regulatory 5'UTRs

The key findings in the second chapter present an expanded model for a novel regulatory system in *D.radiodurans*. The RDR regulatory system has been discovered in recent years and thought to regulate radiation response genes at the promoter level. Our primary evidence for asserting regulation at the 5'UTR for gyrase A was a GMSA testing for binding of DdrO to the mRNA. Moreover, this experiment was supported by the prior finding that the RDRM binding sequence was present in the native mRNA, according to 5' RACE data collected. However, to fortify the claim that DdrO is in fact functioning at the RNA level, I still need to perform certain experiments to confidently defend my assertions.

One critical piece of data I have yet to produce is 5' RACE data confirming the presence of the RDRM in the gyrase A 5' UTR within the fluorescent reporter system. This finding is necessary to confirm that the gyrase A 5' UTR is the same for native transcription and the experimental system developed. I believe the data I currently have suggests that we are likely transcribing the RDRM in the 5'UTR from the pRadGro-GFP reporter system. Additionally, while I am optimistic about the GMSA data, I would like to repeat this rather difficult assay to generate a cleaner image. My aim in redoing this experiment is also to be able to derive affinity data (Kd) for DdrO binding to RDRMs in DNA and RNA, which have yet to be presented. The full characterization of the nature of

DdrO binding to the gyrase A 5' UTR may also serve to advance the theory that the RDR system is acting on UTRs to regulate responses. Thus, it might also be helpful to experimentally verify where the DdrO protein is binding on the gyrase A mRNA. This could be elucidated by performing a DMS foot printing assay to verify the RDRM is, in fact, the binding site for DdrO Finally, the failed genetic analysis presented may also be repeated to produce true gyrase A 5' UTR knockouts for screening. In concert, these additional experiments should solidify the central findings of this work.

Once the aforementioned experiments are conducted, they may confirm the RNA-level functionality of the RDR system for the gyrase A 5' UTR. Subsequent work may seek to characterize other prominent response genes within the RDR to test for 5' UTR regulation. It is possible that the gyrase A 5'UTR is not the only locus of DdrO regulation for the RDR system. The findings presented in this chapter may just be the first step in revealing 5' UTRs' complicity in regulatory systems within *Deinococcus*.

Rational Engineering of Regulatory ncRNA

The *csrB* engineering project is still in development as well. Thus far, I have presented a methodology for assaying accessibility for predicting protein binding-sites for RNA and a methodology for mutating RNA in a modular fashion that conserves structure. As a result, I have ultimately generated a few iterations of *csrB* mutants that exhibit a gradient of Csr regulatory capacity. However, the results and direction of this work can be significantly expanded.

Primarily, the collection of mutants can be vastly expanded to include more *csrB* mutations, and permutations thereof. Testing more *csrB* mutants may reveal more drastic changes in regulatory capacity and yield a comprehensive collection of mutants that may increase the limits of functionality for engineering purposes. Similarly, a final piece of

data to provide this work more completion would be a phenotypic assay demonstrating the "tuning" of gene expression that has only been experimentally determined via the fluorescence reporter system. This assay would validate the application-based value of this bioengineering methodology. Moreover, while the current reporter system (83) has been yielding results, the entire architecture could also be revisited and optimized to possibly yield more accurate results. Instead of using a the current dual plasmid system, the basic machinery used for testing the Csr regulatory capacity via fluorescent could be cloned into different vectors that may have higher copy numbers and different promoters and tested for increased consistency and fidelity.

As this work is in its' relative infancy, there may be many future avenues of investigation that advance the methodology or increase the utility within the Csr system in *E. coli*. Ultimately, the results of this completed work may be applied to other organisms harboring the Csr system and adapted for myriad purposes.

References

1. **Wang HH**, **Kim H**, **Cong L**, **Jeong J**, **Bang D**, **Church GM**. 2012. Genome-scale promoter engineering by coselection MAGE. Nat. Methods **9**:591–3.

2. **Haimovich AD**, **Muir P**, **Isaacs FJ**. 2015. Genomes by design. Nat. Rev. Genet. **16**:501–516.

3. **Stagljar I**. 2016. The power of OMICS. Biochem. Biophys. Res. Commun.

4. **Sandoval NR**, **Kim JYH**, **Glebes TY**, **Reeder PJ**, **Aucoin HR**, **Warner JR**, **Gill RT**. 2012. Strategy for directing combinatorial genome engineering in Escherichia coli. Proc. Natl. Acad. Sci. **109**:10540–10545.

5. **Wiedenheft B**, **Sternberg SH**, **Doudna J a.** 2012. RNA-guided genetic silencing systems in bacteria and archaea. Nature **482**:331–338.

6. **He L**, **Hannon GJ**. 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. **5**:522–531.

7. **Vogel J**, **Luisi BF**. 2011. Hfq and its constellation of RNA. Nat. Rev. Microbiol. **9**:578–589.

8. **Waters LS**, **Storz G**. 2009. Regulatory RNAs in Bacteria. Cell **136**:615–628.

9. **Sorek R**, **Cossart P**. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat. Rev. Genet. **11**:9–16.

10. **Papenfort K**, **Vogel J**. 2010. Regulatory RNA in bacterial pathogens. Cell Host Microbe **8**:116–127.

11. **Winkler WC**, **Breaker RR**. 2005. Regulation of Bacterial Gene Expression By Riboswitches. Annu. Rev. Microbiol. **59**:487–517.

12. **Vakulskas CA**, **Potts AH**, **Babitzke P**, **Ahmer BMM**, **Romeo T**. 2015. Regulation of bacterial virulence by Csr (Rsm) systems. Microbiol Mol Biol Rev **79**:193–224.

13. **Si T**, **HamediRad M**, **Zhao H**. 2015. Regulatory RNA-assisted genome engineering in microorganisms. Curr. Opin. Biotechnol. **36**:85–90.

14. **Wang Z**, **Cirino PC**. 2016. New and improved tools and methods for enhanced biosynthesis of natural products in microorganisms. Curr. Opin. Biotechnol. **42**:159–168.

15. **McKee AE**, **Rutherford BJ**, **Chivian DC**, **Baidoo EK**, **Juminaga D**, **Kuo D**, **Benke PI**, **Dietrich J a**, **Ma SM**, **Arkin AP**, **Petzold CJ**, **Adams PD**, **Keasling JD**, **Chhabra SR**. 2012. Manipulation of the carbon storage regulator system for metabolite remodeling and biofuel production in Escherichia coli. Microb. Cell Fact. **11**:79.

16. **Winkler JD**, **Erickson K**, **Choudhury A**, **Halweg-edwards AL**, **Gill RT**. 2015. Complex systems in metabolic engineering. Curr. Opin. Biotechnol. **36**:107–114.

17. **Makarova KS**, **Omelchenko M V.**, **Gaidamakova EK**, **Matrosova VY**, **Vasilenko A**, **Zhai M**, **Lapidus A**, **Copeland A**, **Kim E**, **Land M**, **Mavromatis K**, **Pitluck S**, **Richardson PM**, **Detter C**, **Brettin T**, **Saunders E**, **Lai B**, **Ravel B**, **Kemner KM**, **Wolf YI**, **Sorokin A**, **Gerasimova A V.**, **Gelfand MS**, **Fredrickson JK**, **Koonin E V.**, **Daly MJ**. 2007. Deinococcus geothermalis: The pool of extreme radiation resistance genes shrinks. PLoS One **2**.

18. **Dubey AK**, **Baker CS**, **Romeo T**, **Babitzke P**. 2005. CsrA − RNA interaction RNA sequence and secondary structure participate in high-affinity CsrA – RNA interaction. Rna 1579–1587.

19. **Giuliodori AM**, **Gualerzi CO**, **Soto S**, **Vila J**, **Tavío MM**. 2007. Review on bacterial stress topics. Ann. N. Y. Acad. Sci. **1113**:95–104.

20. **Butala M**, **Žgur-Bertok D**, **Busby SJW**. 2009. The bacterial LexA transcriptional repressor. Cell. Mol. Life Sci. **66**:82–93.

21. **Marles-Wright J**, **Lewis RJ**. 2007. Stress responses of bacteria. Curr. Opin. Struct. Biol. **17**:755–760.

22. **Huang S**, **Chen L**, **Te R**, **Qiao J**, **Wang J**, **Zhang W**. 2013. Complementary iTRAQ proteomics and RNA-seq transcriptomics reveal multiple levels of regulation in response to nitrogen starvation in Synechocystis sp. PCC 6803. Mol. Biosyst. **9**:2565–74.

23. **Huerta JM**, **Aguilar I**, **López-Pliego L**, **Fuentes-Ramírez LE**, **Castañeda M**. 2016. The Role of the ncRNA RgsA in the Oxidative Stress Response and Biofilm Formation in Azotobacter vinelandii. Curr. Microbiol. **72**:671–679.

24. **Paul D**. 2013. Osmotic stress adaptations in rhizobacteria. J. Basic Microbiol. **53**:101–110.

25. **Qiao J**, **Huang S**, **Te R**, **Wang J**, **Chen L**, **Zhang W**. 2013. Integrated proteomic and transcriptomic analysis reveals novel genes and regulatory mechanisms involved in salt stress responses in Synechocystis sp. PCC 6803. Appl. Microbiol.

Biotechnol. **97**:8253–8264.

26. **Yu W-B**, **Ye B-C**. 2016. Transcriptional Profiling Analysis of Bacillus subtilis in Response to High Levels of Fe3+. Curr. Microbiol. **72**:653–662.

27. **Munteanu A-C**, **Uivarosi V**, **Andries A**. 2015. Recent progress in understanding the molecular mechanisms of radioresistance in Deinococcus bacteria. Extremophiles 707–719.

28. **Slade D**, **Radman M**. 2011. Oxidative Stress Resistance in Deinococcus radioduransMicrobiology and Molecular Biology Reviews.

29. **Krisko A**, **Radman M**. 2013. Biology of Extreme Radiation Resistance. Cold Spring Harb. Perspect. Biol. 1–12.

30. **HARSOJO**, **KITAYAMA S**, **MATSUYAMA A**. 1981. Genome Multiplicity and Radiation Resistance in Micrococcus radiodurans. J. Biochem. **90**:877–880.

31. **Daly MJ**, **Gaidamakova EK**, **Matrosova VY**, **Kiang JG**, **Fukumoto R**, **Lee DY**, **Wehr NB**, **Viteri GA**, **Berlett BS**, **Levine RL**. 2010. Small-molecule antioxidant proteome-shields in Deinococcus radiodurans. PLoS One **5**:10–15.

32. **Tian B**, **Xu Z**, **Sun Z**, **Lin J**, **Hua Y**. 2007. Evaluation of the antioxidant effects of carotenoids from Deinococcus radiodurans through targeted mutagenesis, chemiluminescence, and DNA damage analyses. Biochim. Biophys. Acta - Gen. Subj. **1770**:902–911.

33. **White O**, **Eisen JA**, **Heidelberg JF**, **Hickey l EK**, **Peterson JD**, **Dodson RJ**, **Haft DH**, **Gwinn ML**, **Nelson WC**, **Richardson DL**, **Moffat KS**, **Qin H**, **Jiang L**, **Pamphile W**, **Crosby M**, **Shen M**, **Vamathevan JJ**, **Lam P**, **McDonald L**, **Utterback T**, **Zalewski C**, **Makarova KS**, **Aravind L**, **Daly MJ**, **Minton KW**, **Fleischmann RD**, **Ketchum KA**, **Nelson KE**, **Salzberg S**, **Smith H 0.**, **Venter JC**, **Fraser CM**. 1999. Genome Sequence of the Radioresistant Bacterium Deinococcus radiodurans R1. Science (80-.). **286**:1571–1577.

34. **Devigne A**, **Ithurbide S**, **Bouthier de la Tour C**, **Passot F**, **Mathieu M**, **Sommer S**, **Servant P**. 2015. DdrO is an essential protein that regulates the radiation desiccation response and the apoptotic-like cell death in the radioresistant Deinococcus radiodurans bacterium. Mol. Microbiol. **96**:1069–1084.

35. **Lin L**, **Dai S**, **Tian B**, **Li T**, **Yu J**, **Liu C**, **Wang L**, **Xu H**, **Zhao Y**, **Hua Y**. 2016. DqsIR quorum sensing-mediated gene regulation of the extremophilic bacterium Deinococcus radiodurans in response to oxidative stress. Mol. Microbiol. **100**:527–541.

36. **Yang P**, **Chen Z**, **Shan Z**, **Ding X**, **Liu L**, **Guo J**. 2014. Effects of FMN riboswitch on antioxidant activity in Deinococcus radiodurans under H 2 O 2 stress. Microbiol. Res. **169**:411–416.

37. **Roth A**, **Nahvi A**, **Lee M**, **Jona I**, **Breaker RR**. 2006. Characteristics of the glmS ribozyme suggest only structural roles for divalent metal ions. RNA **12**:607–619.

38. **Tsai C-H**, **Liao R**, **Chou B**, **Contreras LM**. 2015. Transcriptional analysis of Deinococcus radiodurans reveals novel small RNAs that are differentially expressed under ionizing radiation. Appl. Environ. Microbiol. **81**:1754–64.

39. **Storz G**, **Vogel J**, **Wassarman KM**. 2011. Regulation by Small RNAs in Bacteria: Expanding Frontiers. Mol. Cell **43**:880–891.

40. **Oliva G**, **Sahr T**, **Buchrieser C**. 2015. Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. FEMS Microbiol. Rev. **39**:331–349.

41. **Romeo T**, **Vakulskas C a.**, **Babitzke P**. 2013. Post-transcriptional regulation on a global scale: Form and function of Csr/Rsm systems. Environ. Microbiol. **15**:313–324.

42. **Wang Y**, **Xu Q**, **Lu H**, **Lin L**, **Wang L**, **Xu H**, **Cui X**, **Zhang H**, **Li T**, **Hua Y**. 2015. Protease activity of PprI facilitates DNA damage response: Mn(2+)-dependence and substrate sequence-specificity of the proteolytic reaction. PLoS One **10**:1–17.

43. **Ludanyi M**, **Blanchard L**, **Dulermo R**, **Brandelet G**, **Bellanger L**, **Pignol D**, **Lemaire D**, **de Groot A**. 2014. Radiation response in Deinococcus deserti: IrrE is a metalloprotease that cleaves repressor protein DdrO. Mol. Microbiol. **94**:434–449.

44. **Markillie LM**, **Varnum SM**, **Hradecky P**, **Wong KK**. 1999. Targeted mutagenesis by duplication insertion in the radioresistant bacterium Deinococcus radiodurans: Radiation sensitivities of catalase (katA) and superoxide dismutase (sodA) mutants. J. Bacteriol. **181**:666–669.

45. **Misra HS**, **Khairnar NP**, **Kota S**, **Shrivastava S**, **Joshi VP**, **Apte SK**. 2006. An exonuclease I-sensitive DNA repair pathway in Deinococcus radiodurans: A major determinant of radiation resistance. Mol. Microbiol. **59**:1308–1316.

46. **Smith MD**, **Lennon E**, **McNeil LB**, **Minton KW**. 1988. Duplication insertion of drug resistance determinants in the radioresistant bacterium Deinococcus radiodurans. J. Bacteriol. **170**:2126–2135.

47. **Will S**, **Joshi T**, **Hofacker IL**, **Stadler PF**, **Backofen R**. 2012. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA **18**:900–14.

48. **Smith C**, **Heyne S**, **Richter AS**, **Will S**, **Backofen R**. 2010. Freiburg RNA Tools: A web server integrating IntaRNA, ExpaRNA and LocARNA. Nucleic Acids Res. **38**:373–377.

49. **Will S**, **Reiche K**, **Hofacker IL**, **Stadler PF**, **Backofen R**. 2007. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput. Biol. **3**:680–691.

50. **Suess B**, **Fink B**, **Berens C**, **Stentz R**, **Hillen W**. 2004. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res. **32**:1610–1614.

51. **Gelderman G**, **Sivakumar A**, **Lipp S**, **Contreras L**. 2015. Adaptation of Tri-molecular fluorescence complementation allows assaying of regulatory Csr RNA-protein interactions in bacteria. Biotechnol. Bioeng. **112**:365–375.

52. **Vasquez KA**, **Hatridge TA**, **Curtis NC**, **Contreras LM**. 2016. Slowing Translation between Protein Domains by Increasing Affinity between mRNAs and the Ribosomal Anti-Shine-Dalgarno Sequence Improves Solubility. ACS Synth. Biol. **5**:133–145.

53. **Shevchenko A**, **Tomas H**, **Havli\[sbreve] J**, **Olsen J V**, **Mann M**. 2007. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. **1**:2856–2860.

54. **Roth A**, **Winkler WC**, **Regulski EE**, **Lee BWK**, **Lim J**, **Jona I**, **Barrick JE**, **Ritwik A**, **Kim JN**, **Welz R**, **Iwata-Reuyl D**, **Breaker RR**. 2007. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat. Struct. Mol. Biol. **14**:308–317.

55. **de la Tour CB**, **Passot FM**, **Toueille M**, **Mirabella B**, **Guérin P**, **Blanchard L**, **Servant P**, **de Groot A**, **Sommer S**, **Armengaud J**. 2013. Comparative proteomics reveals key proteins recruited at the nucleoid of Deinococcus after irradiation-induced DNA damage. Proteomics **13**:3457–3469.

56. **Basu B**, **Apte SK**. 2012. Gamma radiation-induced proteome of Deinococcus radiodurans primarily targets DNA repair and oxidative stress alleviation. Mol. Cell. Proteomics **11**:M111.011734.

57. **Zhang C**, **Wei J**, **Zheng Z**, **Ying N**, **Sheng D**, **Hua Y**. 2005. Proteomic analysis

of Deinococcus radiodurans recovering from γ-irradiation. Proteomics **5**:138–143.

58. **Luan H**, **Meng N**, **Fu J**, **Chen X**, **Xu X**, **Feng Q**, **Jiang H**, **Dai J**, **Yuan X**, **Lu Y**, **Roberts AA**, **Luo X**, **Chen M**, **Xu S**, **Li J**, **Hamilton CJ**, **Fang C**, **Wang J**. 2014. Genome-wide transcriptome and antioxidant analyses on gamma-irradiated phases of Deinococcus radiodurans R1. PLoS One **9**.

59. **Lu H**, **Gao G**, **Xu G**, **Fan L**, **Yin L**, **Shen B**, **Hua Y**. 2009. Deinococcus radiodurans PprI switches on DNA damage response and cellular survival networks after radiation damage. Mol. Cell. Proteomics **8**:481–494.

60. **Mellin JR**, **Cossart P**. 2015. Unexpected versatility in bacterial riboswitches. Trends Genet. **31**:150–156.

61. **McCown PJ**, **Roth A**, **Breaker RR**. 2011. An expanded collection and refined consensus model of glmS ribozymes. Rna **17**:728–736.

62. **Ren A**, **Rajashankar KR**, **Patel DJ**. 2015. Global RNA Fold and Molecular Recognition for a pfl Riboswitch Bound to ZMP, a Master Regulator of One-Carbon Metabolism. Structure **23**:1375–81.

63. **Meima R**, **Lidstrom ME**. 2000. Characterization of the minimal replicon of a cryptic Deinococcus radiodurans SARK plasmid and development of versatile Escherichia coli-D. radiodurans shuttle vectors. Appl. Environ. Microbiol. **66**:3856–3867.

64. **Anaganti N**, **Basu B**, **Apte SK**. 2016. In situ real-time evaluation of radiation-responsive promoters in the extremely radioresistant microbe Deinococcus radiodurans. J. Biosci. **41**:193–203.

65. **Mercante J**, **Edwards AN**, **Dubey AK**, **Babitzke P**, **Romeo T**. 2009. Molecular Geometry of CsrA (RsmA) Binding to RNA and Its Implications for Regulated Expression. J. Mol. Biol. **392**:511–528.

66. **Earl AM**, **Mohundro MM**, **Mian IS**, **Battista JR**. 2002. The IrrE Protein of Deinococcus radiodurans R1 Is a Novel Regulator of recA Expression. Microbiology **184**:6216–6224.

67. **Teichmann M**, **Dumay-Odelot H**, **Fribourg S**. 2012. Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes **3**:2–7.

68. **De Groot A**, **Roche D**, **Fernandez B**, **Ludanyi M**, **Cruveiller S**, **Pignol D**, **Vallenet D**, **Armengaud J**, **Blanchard L**. 2014. RNA sequencing and proteogenomics reveal the importance of leaderless mrnas in the radiation-tolerant

bacterium deinococcus deserti. Genome Biol. Evol. **6**:932–948.

69. **Makarova KS**, **Aravind L**, **Wolf YI**, **Tatusov RL**, **Minton KW**, **Koonin Eugene V. A**, **Daly MJ**. 2001. Genome of the Extremely Radiation-Resistant Bacterium. Microbiol. Mol. Biol. Rev. **65**:44–79.

70. **Tanaka M**, **Earl AM**, **Howell HA**, **Park MJ**, **Eisen JA**, **Peterson SN**, **Battista JR**. 2004. Analysis of Deinococcus radiodurans's transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. Genetics **168**:21–33.

71. **Joshi B**, **Schmid R**, **Altendorf K**, **Apte SK**. 2004. Protein recycling is a major component of post-irradiation recovery in Deinococcus radiodurans strain R1. Biochem. Biophys. Res. Commun. **320**:1112–1117.

72. **Servant P**, **Jolivet E**, **Bentchikou E**, **Mennecier S**, **Bailone A**, **Sommer S**. 2007. The ClpPX protease is required for radioresistance and regulates cell division after γ-irradiation in Deinococcus radiodurans. Mol. Microbiol. **66**:1231–1239.

73. **Abreu IA**, **Hearn A**, **An H**, **Nick HS**, **Silverman DN**, **Cabelli DE**. 2008. The kinetic mechanism of manganese-containing superoxide dismutase from Deinococcus radiodurans: A specialized enzyme for the elimination of high superoxide concentrations. Biochemistry **47**:2350–2356.

74. **Vazquez-Anderson J**, **Contreras LM**. 2012. Regulatory RNAs Charming gene management styles for synthetic biology applications. RNA Biol. **6286**:3–22.

75. **Gottesman S**, **Storz G**. 2011. Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. Cold Spring Harb. Perspect. Biol. **3**.

76. **Caldelari I**, **Chao Y**, **Romby P**, **Vogel J**. 2013. RNA-mediated regulation in pathogenic bacteria. Cold Spring Harb. Perspect. Biol.

77. **Papenfort K**, **Vanderpool CK**. 2015. Target activation by regulatory RNAs in bacteria. FEMS Microbiol. Rev. 362–378.

78. **Liu MY**, **Gui G**, **Wei B**, **Iii JFP**, **Oakford L**, **Giedroc DP**, **Romeo T**. 1997. The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagonizes Its Activity in Escherichia coli The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagoniz. J. Biol. Chem. **272**:17502–17510.

79. **Timmermans J**, **Van Melderen L**. 2010. Post-transcriptional global regulation by CsrA in bacteria. Cell. Mol. Life Sci. **67**:2897–2908.

80. **Edwards AN**, **Patterson-Fortin LM**, **Vakulskas C a.**, **Mercante JW**, **Potrykus**

K, **Vinella D**, **Camacho MI**, **Fields J a.**, **Thompson S a.**, **Georgellis D**, **Cashel M**, **Babitzke P**, **Romeo T**. 2011. Circuitry linking the Csr and stringent response global regulatory systems. Mol. Microbiol. **80**:1561–1580.

81. **Duss O**, **Michel E**, **Yulikov M**, **Schubert M**, **Jeschke G**, **Allain FH-T**. 2014. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. Nature **509**:588–92.

82. **Sowa SW**, **Vazquez-Anderson J**, **Clark C a.**, **De La Pena R**, **Dunn K**, **Fung EK**, **Khoury MJ**, **Contreras LM**. 2014. Exploiting post-transcriptional regulation to probe RNA structures in vivo via fluorescence. Nucleic Acids Res. **43**:e13–e13.

83. **Adamson DN**, **Lim HN**. 2013. Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation. Proc. Natl. Acad. Sci. **110**:13120–13125.

84. **Gibson DG**, **Young L**, **Chuang R-Y**, **Venter JC**, **Hutchison CA**, **Smith HO**. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth **6**:343–345.

85. **Gudapaty S**, **Suzuki K**, **Wang X**, **Romeo T**, **Wang XIN**, **Babitzke P**. 2001. Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli **183**:6017–6027.

86. **Zuker M**. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31**:3406–3415.

87. **Baker CS**, **Morozov I**, **Suzuki K**, **Romeo T**, **Babitzke P**. 2002. CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli. Mol. Microbiol. **44**:1599–1610.

88. **Duss O**, **Michel E**, **Dit Konté ND**, **Schubert M**, **Allain FHT**. 2014. Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. Nucleic Acids Res. **42**:5332–5346.

89. **Liu MY**, **Gui G**, **Wei B**, **Iii JFP**, **Oakford L**, **Giedroc DP**, **Romeo T**. 1997. The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagonizes Its Activity in Escherichia coli The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagoniz **272**:17502–17510.

90. **Lapouge K**, **Perozzo R**, **Iwaszkiewicz J**, **Bertelli C**, **Zoete V**, **Michielin O**, **Scapozza L**, **Haas D**. 2013. RNA pentaloop structures as effective targets of

regulators belonging to the RsmA/CsrA protein family. RNA Biol. **10**:1031–41.

Vita

Paul Amador was born in Richmond, Texas. Upon graduation from Morton Ranch High School in May 2010, he attended the University of the Incarnate Word in San Antonio, Texas and received a Bachelor of Science degree in biology in May 2013. Thereafter, he started graduate school at the University of Texas at Austin in June 2013.

Email Address: paulamadorable@gmail.com

This thesis was typed by the author.